

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2003, 15:03:05 ; Search time 1969 seconds  
(without alignments)  
11022.438 Million cell updates/sec

Title: US-09-758-269-5  
Perfect score: 1800  
Sequence: 1 atggctttttcacggcaac.....tggcgaagcaggtgtgga 1800

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*

- 1: em\_estba.\*
- 2: em\_esthum.\*
- 3: em\_estin.\*
- 4: em\_estmu.\*
- 5: em\_estov.\*
- 6: em\_estpl.\*
- 7: em\_estro.\*
- 8: em\_htc.\*
- 9: gb\_estl.\*
- 10: gb\_est2.\*
- 11: gb\_htc.\*
- 12: gb\_est3.\*
- 13: gb\_est4.\*
- 14: gb\_est5.\*
- 15: em\_estfun.\*
- 16: em\_estom.\*
- 17: em\_gss\_hum.\*
- 18: em\_gss\_inv.\*
- 19: em\_gss\_pln.\*
- 20: em\_gss\_vrt.\*
- 21: em\_gss\_fun.\*
- 22: em\_gss\_mam.\*
- 23: em\_gss\_mus.\*
- 24: em\_gss\_pro.\*
- 25: em\_gss\_rod.\*
- 26: em\_gss\_phg.\*
- 27: em\_gss\_vrl.\*
- 28: gb\_gss1.\*
- 29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	589.8	32.8	837	28	BH62445	BH62445 BOHTU09TR
2	504	28.0	627	9	AV826228	AV826228 AV826228
3	459.6	25.5	766	29	BZ457573	BZ457573 BONGH80NF
4	425.6	23.6	720	12	BM412731	BM412731 EST567047

5	417	23.2	805	12	BM408615	BM408615 EST582942
6	416.2	23.1	781	28	BH549344	BH549344 BOGR053TF
7	412.4	22.9	721	28	BH739063	BH739063 BOMHT72TF
8	393.8	21.9	644	12	BM535408	BM535408 EST58430
9	382.8	21.3	643	10	AW933524	AW933524 EST39283
10	375.6	20.9	787	12	BM408565	BM408565 EST582892
11	356.2	19.8	667	29	AL950790	AL950790 Arabidops
12	344	19.1	592	10	BF113346	BF113346 EST440936
13	334.2	18.6	695	28	BZ022957	BZ022957 oeh32c12
14	324.4	18.5	565	10	BE451924	BE451924 EST413439
15	326	18.1	637	13	BUS50566	BUS50566 GM880021A
16	324.2	18.0	562	12	BM536135	BM536135 EST589157
17	323.2	18.0	553	12	BM085672	BM085672 saJ28a02
18	318	17.7	559	12	BM085005	BM085005 saJ31a08
19	316.6	17.6	520	10	BE432853	BE432853 EST399478
20	313.8	17.4	549	12	BM891057	BM891057 sam22d05
21	313.8	17.4	742	28	BH559796	BH559796 BOHR347TF
22	312.8	17.4	594	28	BH458011	BH458011 BOHBT02TF
23	309.4	17.2	547	12	BE1974879	BE1974879 sai74b11
24	308.6	17.1	547	10	BE437072	BE437072 EST408190
25	304.8	16.9	578	29	EX004482	EX004482 Arabidops
26	303.4	16.9	617	10	BE458861	BE458861 EST414153
27	300.4	16.7	545	12	BM084948	BM084948 saJ30b08
28	296.8	16.5	532	12	BJ563195	BJ563195 BJ563195
29	295.8	16.4	618	9	AW443298	AW443298 EST308228
30	295.8	16.4	711	29	BZ424431	BZ424431 BONDRI4TR
31	294	16.3	564	12	BM085488	BM085488 saJ37d09
32	292.8	16.3	627	10	BF051297	BF051297 EST436472
33	289.2	16.1	509	10	BE451573	BE451573 EST402461
34	288.8	16.0	495	10	BE434930	BE434930 EST406008
35	285.2	15.8	649	29	CC016051	CC016051 PUDGL12TD
36	282.8	15.7	567	13	BO582721	BO582721 E012280-0
37	281.4	15.6	651	29	BZ462969	BZ462969 BONDN77TR
38	277.8	15.4	641	29	BZ535251	BZ535251 OGALK24TT
39	276.2	15.3	490	10	BF050563	BF050563 EST435721
40	275.4	15.3	657	9	AA556214	AA556214 69 lOb101
41	267.4	14.9	503	28	BZ7476	BZ7476 T9A21TR TAM
42	267.4	14.9	602	13	BQ505126	BQ505126 EST612541
43	265.8	14.8	668	29	CC157378	CC157378 ig19b12.b
44	262.8	14.6	612	29	BZ525247	BZ525247 OGALK24TC
45	254.8	14.2	566	10	BE459895	BE459895 EST415187

ALIGNMENTS

RESULT 1  
BH62445/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BH62445 837 bp DNA linear GSS 19-FEB-2002  
BOHTU09TR BO\_2\_3 KB Brassica oleracea genomic clone BOHTU09,  
genomic survey sequence.  
BH62445 GI:18721309  
GSS.  
Brassica oleracea  
Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; eurosids II; Brassicales; Brassicaceae; Brassica.  
1 (bases 1 to 837)  
Town,C.D., Van Aken,S., Uterback,T., Koo,H. and Fraser,C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished  
Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR  
Class: sheared ends.  
Location/Qualifiers

FEATURES

```
RESULT 2
LOCUS AV826228
DEFINITION AV826228 RAFL8 Arabidopsis thaliana cDNA clone RAFL08-11-H16 5',
mRNA sequence.
ACCESSION AV826228
VERSION AV826228
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 627)
AUTHORS Seki, M., Nakamura, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, Y.,
Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.
and Shinozaki, K.
TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: msek@rtc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified lambda PLC-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified paluscript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
details.
FEATURES
source Location/Qualifiers
1..627
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL08-11-H16"
/dev_stage="rosette plants"
/lab_host="DH10B"
/clone_lib="RAFL8"
/notes="Site 1: BamHI; Site 2: SalI; subjected to
dehydration-treated (1, 2, 5, 10, 24 hr)"
BASE COUNT 162 a 179 c 128 g 158 t
ORIGIN
Query Match 28.0%; Score 504; DB 9; Length 627;
Best Local Similarity 100.0%; Pred. No. 1.3e-142;
Matches 504; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCTTCTTTCACGGCAACGGCTGCGGTTCTTGGGAGATGGCTTGGTGGCAATCATACT 60
DB 124 ATGGCTTCTTTCACGGCAACGGCTGCGGTTCTTGGGAGATGGCTTGGTGGCAATCATACT 183
QY 61 CAGCGGCCATTATCGTCTTCTCAAGCTCCGACTTGAAGTTATTTAGTCTCTTACCTATG 120
DB 184 CAGCGGCCATTATCGTCTTCTCAAGCTCCGACTTGAAGTTATTTAGTCTCTTACCTATG 243
QY 121 GCGAGTCGTCTCACACCTAGCTCAATGTTTCTATCGCTTCCACTCTCCAGCTCTT 180
DB 244 GCGAGTCGTCTCACACCTAGCTCAATGTTTCTATCGCTTCCACTCTCCAGCTCTT 303
QY 181 CATTTCCCTTAAGCAATCATCAAACTCTCCGCCATTGTTTAAAGCCAAAGCAAGAA 240
DB 304 CATTTCCCTTAAGCAATCATCAAACTCTCCGCCATTGTTTAAAGCCAAAGCAAGAA 363
QY 241 TCCACACTAAACAGATGATTTGTTCCAGAGCGCGCGGAGGTTGGACGGCGG 300
DB 364 TCCACACTAAACAGATGATTTGTTCCAGAGCGCGCGGAGGTTGGACGGCGG 423
QY 301 GAGGGTTTCTTGTGTGACGCCAGAGACTACCCGCTTCTTAAACGGCTGATCTT 360
```

```

424 GAGGTTTCTCTGTCAGCCAGAGACTACCCGTTTCTTAAACGGTGTATCTCTAGT 483
425
426 GTTCAGATCGCGGAAATTTTCTCCGGTGAATGAACAGCCCGTCCGGCGTAAATCTTCCG 420
427
428 GTTCAGATCGCGGAAATTTTCTCCGGTGAATGAACAGCCCGTCCGGCGTAAATCTTCCG 543
429
430 GTGTCGGAAATCTCCGATTCATCAAGAGAGTGTATGTGGCAACGGAGCTAACCCA 480
431
432 GTGTCGGAAATCTCCGATTCATCAAGAGAGTGTATGTGGCAACGGAGCTAACCCA 603
433
434 CTTACAGAGCGGTGACAGGTTCAC 504
435
436 CTTACAGAGCGGTGACAGGTTCAC 527
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

```

RESULT 3
LOCUS BZ457573/5
DEFINITION BONGT80TF BO.1.6.2 KB tot Brassica oleracea genomic clone BONGT80,
genomic survey sequence.
ACCESSION BZ457573
VERSION BZ457573.1 GI:26737071
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 766)
Town,C.D., Van Aken,S., Uterback,T., Koo,H. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished
Other GSSs: BONGT80FR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1..766
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BONGT80"
/clone_lib="BO.1.6.2_KB_tot"
/note="Vector: pHD1; Site 1: BstXI; 1.6-2 kb sheared
total DNA inserted into pHD1 using BstXI linkers"
BASE COUNT 177 a 203 c 203 g 183 t
ORIGIN
Query Match 25.5%; Score 459.6; DB 29; Length 766;
Best Local Similarity 76.7%; Pred. No. 6.1e-129;
Matches 575; Conservative 0; Mismatches 174; Indels 1; Gaps 1;
QY 529 GTTCACGCGGTCAATTCGAACACGGTTCAGTACGCTTGGCGGTTCCTACTCAGACT 588
DB 751 GATCAGCGGTCCGGTTTAAATACCGTTCGGTTAGTTATTCGTGCGGTTCACGAACA 692
QY 589 A-ACGGTTTGTTCAGGAACGTCATNTGGGTGCGCGGTTCCTCCCAAGCCATCGGTGA 647
DB 691 ACCCGGTTTGTTCAGAGCGAGAGTCGGCGGTCTGTTTCCCAAGCGATAGGAGA 632
QY 648 GTTCACGCGCACACCGGTATTGCGGACTCATGTTCTACGCCAGCTGCGACGCGG 707
DB 631 GTTCACGAGACTTGGGATCGCTAAGCTTAAGTATTAACGACGCTGGCGTTTCGG 572
QY 708 TATAGTCGACCGCGCACACCGGTTAGCTAACCGCGGTTCGTCTATTTCAATGG 767

```

```

571 TTTACTCGACCCGACCCAGGACTCGGTGTGGCTAATCCCGTTTAGTCTATTTCACCG 512
572
573 COGTTTATTGGCTATGTTCGGAGGATGATTACCTTTACCAAGTTCAGATCCTCCCAATGG 827
574
575 TCATCTCTTAGCAATGTTCGGAAGACGATTACCGGTACCACTAAAGTTCACCAACCGG 452
576
577 AGATTAAACACCGTTGTCGGTTTCGATTTCGATTCGACAAATAGTAATCACAATGATTCG 887
578
579 AGATTAAACACCGTTGTCGGTTTCGATTTCGATTCGACAAATAGTAATCACAATGATTCG 392
580
581 CCACCCGAAAGTCGACCCGGAATCCGGTGAATCTTCGCTTTAAAGTACGACGCTGTTTC 947
582
583 CCACCCGAAAGTCGACCCGGAATCCGGTGAATCTTCGCTTTAAAGTACGACGCTGTTTC 332
584
585 AAAGCTTACCTAAATATCTTCGATTTCACCCGAGGAACTTAATCACCAGGCTCGA 1007
586
587 GAAGCTTACCTAAATATCTTCGATTTCGATTCGATTCGATTCGATTCGATTCGATTCG 272
588
589 GATTGAGTTGATTCAGCAACGATGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 1067
590
591 GATCCCGTTCGACCAACCGAGATGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 212
592
593 CGTACCTGACCAAGTTCGTTTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 1127
594
595 GGTTCGCGACCAAGTTCGTTTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 152
596
597 GGTTCGCGACCAAGTTCGTTTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 1187
598
599 GGTTCGCGACCAAGTTCGTTTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 92
600
601 ATCGAACATTAAGTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 1247
602
603 TTCTGCTGATCGGTGGATCGAGTGGAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 32
604
605 GGAAGAGCCAGAAACAGATGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 1277
606
607 GGAAGAGCCAGAAACAGATGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 2
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

## RESULT 4

BM412731

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1..720

/organism="Lycopersicon esculentum"

/mol\_type="mRNA"

/cultivar="TA496"

/db\_xref="taxon:4081"

Seq primer: T3.

Contact: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

This clone is available through the Clemson University Genomics

Institute

Generation of ESTs from tomato fruit tissue, breaker stage (2002)

Unpublished

EST

EST

EST

EST

EST

```

/clone="cLEG50N24"
/tissue_type="Pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/clone_lib="tomato breaker fruit"
/notes="Vector: pBluescriptKMCuadapt; Site 1: EcoRI;
Site 2: XhoI; supplier: Boyce Thompson Institute;
sequencing: The Institute for Genomic Research. Fruit
were harvested at the breaker stage (first sign of
lycopene accumulation on the blossom end of fruit). Fruit
were cut in half and the seeds and locules were discarded
prior to freezing the pericarp."
BASE COUNT      226 a 143 c 155 g 196 t
ORIGIN

Query Match      23.6%; Score 425.6; DB 12; Length 720;
Best Local Similarity 74.4%; Pred. No. 1.5e-118;
Matches 536; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

QY 740 CTAACGCCGTTGGTCTATTTCATGGCGGTTATTGCTATGTCGGAGGATGATTAC 799
DB 1 CAACACGCCGTTAGTCTATTTCATACCGATTCTTGCTATGCTGAAGATGATTGC 60
QY 800 CTTACCAAGTTTCAGATCACTCCCAATGGAGATTAAACACCGTTGGTTCGATTTC 859
DB 61 CTTACCATGTAAGAGTAACACCCACCGCGGATCTTAAACACAGAGGTCGATTTCG 120
QY 860 ATGGACAATTAGAATCCACATGATGCCACCCGAAAGTCGACCGGAATCCGTTGAAC 919
DB 121 ACGCCAGCTAAATCCACCATGATAGTACCACCAAGCTGACCCAGTTCCGGTGAGC 180
QY 920 TCTTCGCTTAAAGTACGAGCTGTTTCAAGCTTTACCTAAATACCTTCGATTCTCAC 979
DB 181 TATTGCTCTTAGCTAGATGTGATTTCAGAAAGCATACCTCAAGTACTTCAGATTTCAA 240
QY 980 CGGACGGAATTAATACCGGAGTGCAGATTTCAGTTCATCAGCCACGATGATGACG 1039
DB 241 ABAATGGGAAATCAATGATGTTGAATTCAGTTGAGACCCACCAATGATGATG 300
QY 1040 ATTTCGCGATTACAGAAATTCGTGCTGCTACTGACCAAGTCGTTTCAAGCTGC 1099
DB 301 ATTTCGGAATTAAGTACGGAAGTTCGTGCTATTCCTGATCAACAAAGTCGTTTCAAGATGT 360
QY 1100 CGGAGATGATCCCGGTTGGTCTCCGTTGTTTACGACCAAGACGATGATGATTCG 1159
DB 361 CTGAATGATCCGTTGGAGTTTCACCGGTGTTTACGACCAAGACAAAGTTCCCGATTTC 420
QY 1160 GGATTTAGACAAATACCGCGAAGATTCATCGAATTTAAGTGGATTGATGCTCCAGATT 1219
DB 421 GTATTCTGGATTAAGTACGGAAGATGGGTCTGATTGAAATGGGTGAAATACCTGATT 480
QY 1220 GCTTCTGCTTCCATCTCTGGAAGCTTGGGAAGAGCCAGAAACAGATGAAGTCGTGTA 1279
DB 481 GTTCTGTTTCCACCTCTGGAATGCTTGGGAAGACGACAGAAATGAAATCGTTGTAA 540
QY 1280 TAGGGTCTGTATGATCCACGATCAATTTTCAACGAGTCTGACGAAATCTTCAGA 1339
DB 541 TTGTTTCATGATGACACCAACGACATCCATTTTCAATGAATGATGATGAAGGCTTAAAGA 600
QY 1340 GTGTCTGCTGTGAATCCGCTGAATCTCAAAACCGGTGAAATCAACTCCCGTCCGATCA 1399
DB 601 GTGTTTATCGAAATCCGTTCTCAATTTTGAACACGGGAATTCACAGAAATCCATAA 660
QY 1400 TCTCCACGAGATCAACAGTCAACCTCGAAGCGGATGTTCAACAGAAATGCTGTCG 1459
DB 661 TCGAAACCCGGATGAACAAAGTGGATTTAGAAGCTGGAATGTTGTAACCCGAAACAAACTCG 720

RESULT 5
BV408615
LOCUS
DEFINITION EST582842 tomato breaker fruit Lycopersicon esculentum cDNA clone
cLEG50N24 5' end, mRNA sequence.

```

```

ACCESSION      BM408615
VERSION        BM408615.1
KEYWORDS       GI:18260245
SOURCE         Lycopersicon esculentum (tomato)
ORGANISM       Lycopersicon esculentum
REFERENCE      1 (bases 1 to 805)
AUTHORS        Alcala J., Vrebalov J., White R., Vision T., Karanymcheva S.A., Tsai
                J., Bougri O., Kirkness E., Utterback T., Van Aken S., Rønning
                C.M., Fraser C.M., Martin G.B., Tanksley S.D. and Giovannoni J.
TITLE          Generation of ESTs from tomato fruit tissue, breaker stage (2002)
JOURNAL        Unpublished
COMMENT        Contact: CUGI
                Clemson University Genomics Institute
                Clemson University
                100 Jordan Hall, Clemson, SC 29634, USA
                Email: http://www.genome.clemson.edu/orders/index.html
                This clone is available through the Clemson University Genomics
                Institute
Seq primer: T3.
Location/Qualifiers
1..805
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEG50N24"
/tissue_type="Pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/clone_lib="tomato breaker fruit"
/notes="Vector: pBluescriptKMCuadapt; Site 1: EcoRI;
Site 2: XhoI; supplier: Boyce Thompson Institute;
sequencing: The Institute for Genomic Research. Fruit
were harvested at the breaker stage (first sign of
lycopene accumulation on the blossom end of fruit). Fruit
were cut in half and the seeds and locules were discarded
prior to freezing the pericarp."
BASE COUNT      262 a 144 c 180 g 219 t
ORIGIN

```

```

Query Match      23.2%; Score 417; DB 12; Length 805;
Best Local Similarity 71.8%; Pred. No. 7.1e-116;
Matches 575; Conservative 0; Mismatches 220; Indels 6; Gaps 2;

QY 899 TCGACCCGGAATCCCGTGAACTCTTCGCTTAAAGTACGAGCTCGTTTCAAGCTTACC 958
DB 6 TCGACCCAGTTTCCCGTGAGCTATTTCCTCTTAGCTAGCTGTGATTTCAGAGCCATACC 65
QY 959 TAAATACTTCGGATTCTCACCGGAGGAATTAATCACCGAGCTCGAGATTTCAGCTTG 1018
DB 66 TCAAGTACTTCAGATTTCAAAAATGGGAAATCAATGATGTTGAATTCAGTTG 125
QY 1019 ATCAGCCACGATGATGACGATTTTCGGGATTACAGAACTTCGTCTGCTACTGAC 1078
DB 126 AAGACCCCAATGATGATGATTTTCGCAATTTACTGAGAACTTCGTCTGCTCATCTG 185
QY 1079 AGCAAGTCGTTTTCAGCTTCGCGGAGATGATCCGCGTGGGTCTCCGTTGGTTTACGACA 1138
DB 186 AACAAGTCGTTTTCAGATGCTCTGAATGATCCGTTGGAGTTTCACCGTGGTTTACGACA 245
QY 1139 AGAACAGGTTCGCAAGATTCGGGATTTAGACAAATACGCCGAGATTCATCGAACATTA 1198
DB 246 AGAACAAAGTTTCCCGATTTCGATTTCTGGATTAAGTACGCCGAAAGATGGGTCTGATTGA 305
QY 1199 AGTGGATTGATGCTCCAGATTGCTTCTGCTTCATCTCTGAAACCTTGGGAAGAGCCAG 1258
DB 306 AATGGGTTGAGTACCTGATTCCTTCTGTTTCCACTCTGGAATGCTTGGGAGAGCAG 365
QY 1259 AACAAGATGAGTCTGCTGATGATGAGGTCCTGATGACTCCACGAGACTCAATTTTCAACG 1318

```



Db 366 AACAGATGAATCGTTGTAATGGTTTCATGTATGACACACCAGACTCCATTTCAATG 425

Qy 1319 AGTCTGACGAAATCTCAAGAGTGTCTGTCTGAAATCCGCTGAATCTCAAAACCGGTG 1378

Db 426 AATGTGATGAGGGTAAAGAGTGTCTTATCCGAAATCCGCTCAATTTGAACACAGGA 485

Qy 1379 AATCAATCGCGCTCGATCATCTCCACGAGAGTCAACAGTCAACCTCGAAGCAGGA 1438

Db 486 AATCAACAAGAAATCCATAATCGAAACCCGGATGAACAGTGAATTTAGAGCTGAA 545

Qy 1439 TGCTCAACAGAAATGCTCGCGGTAAACCAATTCGCTTACTTGGCTTTAGCCGAGC 1498

Db 546 TGGTGAACCCGAAACAAATCCGAAAGGAAACAGAGTATGCTTATTTGGCTATCGCTGAAC 605

Qy 1499 CGTGCCTTAAAGTCTCAGGATTCGCTAAAGTGTATCTCACTACTGGAAGATTAAAGAAC 1558

Db 606 CATGCCAAAAGTTTCTGGTTTGCAAAAGTAAACCTGTTCAACCGGTGAAGTTGAGAAAT 665

Qy 1559 ATCTTTACGGGATACCGTTACGGAGAGAGCTCTGTTCTCCCGGAGAGGAGGA- 1617

Db 666 TCATTTATGGTGACAAACAAATATGTTGGGGAACCTCTTTTTTAAAGAGAACCAACAG 725

Qy 1618 ----GAGGAACAGGAAGATACATCTCTGTTTCGTTCAACGAGAGAGACATGGAATC 1673

Db 726 CCAGGAAGAGACGATGTTATTTTATTTTACCTTCTGTTCAAGATGAGAA-AGATGGGAATC 784

Qy 1674 GGAGTTACAGATAGTTAAAGC 1694

Db 785 AAAACTGCAATTTGGTAAGC 805

## RESULT 6

BH549344

LOCUS BOGRQ53TF BOGR Brassica oleracea genomic clone BOGRQ53, genomic survey sequence.

ACCESSION BH549344

VERSION BH549344.1 GI:17801124

KEYWORDS GSS.

SOURCE Brassica oleracea

ORGANISM Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 781)

AUTHORS Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.

TITLE Whole genome shotgun sequencing of Brassica oleracea

JOURNAL Unpublished

COMMENT Other\_GSSs: BOGRQ53TR

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

## FEATURES

## source

1..781

/organism="Brassica oleracea"

/mol\_type="genomic DNA"

/strain="TO1000DH3"

/db\_xref="taxon:3712"

/clone="BOGRQ53"

/clone\_lib="BOGR"

/note="Vector: pHS01; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHS01 using BstXI linkers"

## BASE COUNT

## ORIGIN

## Query Match

Best Local Similarity 70.8%; Pred. No. 1.2e-115; Length 781;

Matches 553; Conservative 0; Mismatches 228; Indels 0; Gaps 0;

Qy 274 GGGGGGGGGGAGCGTTTGACCGCGCGAGGGGTTTCTTGTTCAGCCAGGAGCTACAC 333

Db 1 GCAGCGCGGATTGCCATCGATGTGCTGCGGAGGGTTGATCTCAGCTGAGCAAGATACT 60

Qy 334 CGCTTCTTAAACCGGCTGATCTCTAGTGTTCAGATCGCGGGAATTTTGTCCGCTGAAT 393

Db 61 CTTCTTCCAAACCGCTGATCCAGTATTCAAATCGCGGGGAATTTATTTCCCGGTGGC 120

Qy 394 GAACAGCCGCTCCGCGTAATCTTCGGTGTGCGGAAAACCTTCCCGAATTCATCAAGGA 453

Db 121 GAATTTCCGCTCGAGAGTGCTTCGACGTTGAAGGAACAATCCCTGACTGCAATTAACGA 180

Qy 454 GTGTATGTGCGCAACCGAGCTTAACCCATTCAAGAGCGGTGACAGGTCACACTTCTTC 513

Db 181 GTTATGTCCGTAACCGGTGCAATCCGATGTTTCGAGCCAAATCGCTGGGCATCATTTGTC 240

Qy 514 GACGAGAGCGGTATGTTTCAACCGCTCAAAATTCGAACACGGTTTCAGCTAGCTACGTTGC 573

Db 241 GACGAGAGCGGAATGGTTCCAGCTTAAATTAACCAACGGTTTCAGCAAGCTACGATGC 300

Qy 574 CGGTTTACTCAAGCTAACCGGTTTTCAGGAACGCTCAATTCGGGTGACCGGTTTCCCG 633

Db 301 CGGTTTACTTAAACCGGAGAGATTAATTTCAAGAAAACGATTCGGGTGACCGGTTTCCG 360

Qy 634 AAAGCATCGGTGAGCTTCAACGCGCACACCGGTATTGCCGACTCATGCTATTTACGCG 693

Db 361 AAAGCAATCGGAGGCTTCAACGCGCACTCGGGAATCGCACGTCGATGCTTTTTACGA 420

Qy 694 AGAGCTGACGCGGTATAGTCGACCGGCAACCGGTAACCGGTGTAGCTAACCGCGTTTG 753

Db 421 CGTGGGCTTTGGGCGCTCGTCAACAATCAAAACCGGTTCGGGTAGCGAACCGCGTTTG 480

Qy 754 GTCATTTTCAATCGCGGTTATGCTATGCTGTCGAGGATGATTTACCTTACCAAGTTCAG 813

Db 481 GTTACTTCAATAACCGGCTTTTAGCAATGTGAGAAAGAGATTTACCGTACCAATTAAGA 540

Qy 814 ATCACTCCCAATGGAGATTTAAAAAACCGTTGGTTCGGTTTCGATTTTATGAGCAATPAGAA 873

Db 541 ATAACCTCAAAACCGGCGACCTCCAAACCGTTGGGCGTTACGATTTCCGCGTCAGTTAAA 600

Qy 874 TCACAAATGATTTCCCAACCGGAAAGTCGACCGGGAATCCGGTGAAGCTTTCGCTTTAAGC 933

Db 601 TCATCATGATCGCCACCGGAAACTCGACCGGTTTACAAAGAGCTCCACCGCTTAAGC 660

Qy 934 TAGCAGTCTGTTTCAAAGCCTTACCTTAAATACTTCCGATTTCTCAGCGACGGAATAAA 993

Db 661 TAGCAGTCTGTTTAAAGGCTTACCTGAATATTTAGATTTCTCGCGGACGCGTTAAA 720

Qy 994 TCACCGGAGCTCGAGATTCAGCTTGTATGATGAGCCAAACGATGATGACGATTTCCGATTACA 1053

Db 721 TCACCGGAGCTCGAGATTCGCGCTCGAGACTCCGAGACTCCGAGATGTTTCAGATTTCCGTTAACG 780

1054 G 1054

781 G 781

## RESULT 7

## BH739063

## LOCUS

## DEFINITION

## ACCSSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## 1 (bases 1 to 721)

## BOMHT72F

## BO\_2\_3\_KB

## Brassica oleracea genomic clone BOMHT72,

## genomic survey sequence.

## BH739063

## BH739063

## GI:18844458

## GSS.

## Brassica oleracea

## Brassica oleracea

## Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Brassicales; Brassicaceae; Brassica.

AUTHORS Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.N.  
 TITLE Whole genome shotgun sequencing of *Brassica oleracea*  
 JOURNAL Unpublished  
 COMMENT Other GSSs: BOMH72TR  
 Contact: Chris Town

TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3323  
Fax: 301-838-0308  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: 1F  
Class: sheathed ends.

[illegible]

```

1. 721
/organism="Brassica
/mol_type="genomic
/strain="TOL000DH3"
/db_xref="taxon:371
/clone="BOMHT72"
/clone_lib="BO_2_3
/note="vector: pHOS
genomic DNA inserte
178 c 198 g

```

BASE COUNT	188 a	178 c	198 g	157 t
ORIGIN				
Query Match		22.9%	Score 412.4;	DB 28; Length 721;
Best Local Similarity		75.2%	Pred. No. 1.7e-114;	
Matches 531: Conservative			0; Mismatches 166;	Indels 9; Gaps 1;

Qy	904	CCGGAAATCCGGTGAACCTCTTCGCTTTTAAGCTACGACGTCGTTTCAAGAGCCTTACCTAAA	963
Db	4	CCGGGAACCCGCGAGCTATTGCTCTCTCAGCTACGACGTCGTTTCAAGAGCCTTACTTAAA	63

Qy	964	TACTTCCGATTCTCACCGGACGGAACTAAATCAACCGGACGTCAGATTCAGCCTTGATCAG	1023
Db	64	TACTTTCAGATTTCGATCCGATCCGACGTCGCGAAATCCCGCGACGTCGAGATCCCGCTCGACCAA	123

QY	1024	CCAAACGATGATGCACGATTTTCGGATTACAGAGAACTTCGTCTGTAACCTGACCAGCAA	1083
DB	124	CCGACGATGATCCACGACTTTCGGATCAGCGAGAACTTCGTCTGTTGTTTCGACCAAGCAA	183

Qy	1084	GTCTTTTCAAGTCCGCGAGATGATCCCGGTGGGTCTCCGGTGGTTTACGACAGAAC	1143
Dβ	184	GTGGTGTTCAGGCTACCGAGAGATGATCCGCGCGCTCTCCGGTGGTTTACGACAAAAC	243

Accession	Sequence	Length
QY	1144 AAGTCGCGAAGATTCGGGAGTTTATAGACAAATA CGCGGAAGATTTCATCGAACATTAACTGG	1203
DB	244 AAGAAATCAAGATTTCGGTGTCTTTTGGAGCAAAATACGCTGAAGACGGTTTCGTCGATCGCGTGG	303

Accession	Sequence	Length
QY 1204	ATTGATGCTCCAGATTGCTTCTGCCTTCCATCTCTGGAAACGCTTGGGAAGACCGCAAAACA	1263
Db 304	ATCAGAGGTGGAAGACTGCCTTCTGTTTCCATCTCTGGAAACGCTTGGGAAGACCGGAGACA	363

Qy	1264	GATGAAGTCGTGCTGATAGGTCCTGCTATGACTCCACAGACTCAATTTTCAACAGAGTCT	1329
Db	364	GACAGAGTCGTGCTGATCCGGTCATGCGTCATGCATGACGCGCCGACCTCGATATTACGAACAC	423

Qy	1324	GACGAGAACTCTCAAGAGTGTCCTCTCTGAAATCGCGCTGANTCTCAAAACCGGTGAATCA	1383
Dδ	424	GACGAAACACTTTCAGAGTGTTTTGTCGGAGATAAGACTAAACCTGAAACACGGGGAGTCC	483

Qy	1384	484	Db
	ACTGCGCGTCCGATCATCTCTCAACGAAGATCA	ACGCGTCCGCGGGTATCTCC-----GAGAAATTAATCTCGAAGCGGTATGGTA	1445
	AGAGTCAACTCTCGAAGCAGGATGGTC		534

1444 AACGAAACATGCTCGGCCGTAAACCAAAATTCGTTACTTGGCTTAGCCGAGCCGTG 1503

535 AACCGGAATCTTTTAGGTAGAAAAACGCGGTTCTGTTACCTGCTTTACCGAACCGTGG 594

Qy Db

1504 CCTAAAGTCTCAGGATTCGGTAAAGTTGATCTCATCTACCTCGGAGAGTTTANGAAATCTT 1564  
595 CCTAAAGTGTCCGGTTTCGGGAAATGGACTTATCTACCGGAGAAAGTTCNAAAATATATA 654

Qy  
1564 TACGGCGATAACCGGTTACGAGGAGAGCCCTCTGTTTCTCCCCGGAG 1609

Db  
655 TACGGTGACGGTAAAGTCGAGGAGAGCCCTCTGGTTATGTCGGAAG 700

RESULT 8	ACCESSION
BM535408	VERSION
LOCUS	KEYWORDS
DEFINITION	SOURCE
	ORGANISM

REFERENCE AUTHORS	TITLE JOURNAL COMMENT
----------------------	-----------------------------

FEATURES  
SOURCE

BASE COUNT  
ORIGIN

Query Match  
Best Local  
Matches

3

古

2

3

Qy

D6

Qy

D'b

```

Contact: Victor
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
Location/Qualifiers
1..643
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLSP54E2"
/tissue_type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOUR"
/clone_lib="tomato fruit mature green, TAMU"
/notes="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2:
XhoI; cDEF - Fruit were tagged at the 1cm stage and
harvested 3-5 days prior to ripening. Fruit were cut in
half to verify the seeds were indeed 'immature' and the
seeds and locules were discarded prior to freezing the
pericarp"

```

ACCESSION	CLEG45E23	5' end, mRNA sequence.
VERSION	EM408565	
KEYWORDS	EM408565.1	GI:18260195
SOURCE	EST.	
ORGANISM	Lycopersicon esculentum (tomato)	
	Lycopersicon esculentum	
	Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon. 1 (bases 1 to 787)	
REFERENCE	1 (bases 1 to 787)	
AUTHORS	Acala, J., Vrebalov, J., White, R., Vision, T., Karamycheva, S.A., Tsai J., Bougri, O., Kirkness, E., Uterback, T., Van Aken, S., Renning, C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.	
TITLE	Generation of ESTs from tomato fruit tissue, breaker stage (2002)	
JOURNAL	Unpublished	
COMMENT	Contact: CUGI	
	Clemson University Genomics Institute	

Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Email: <http://www.genome.clemson.edu/orders/index.html>  
This clone is available through the Clemson University Genomics Institute  
Seq primer: T3.  
Location/Qualifiers  
1..787  
/organism="Lycopersicon esculentum"  
/mol\_type="mRNA"  
/cultivar="TA496"  
/db\_xref="taxon:4081"  
/clone="CLEG4523"  
/tissue\_type="pericarp"  
/dev\_stage="breaker"  
/lab\_host="SOLR"  
/clone\_lib="tomato breaker fruit"  
/note="Vector: pBluescriptSMCudapt; Site1: EcoRI; Site2: XhoI; supplier: Boyce Thompson Institute; sequencing: The Institute for Genomic Research. Fruit were harvested at the breaker stage (first sign of lycopene accumulation on the blossom end of fruit). Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp."  
BASE COUNT 216 a 194 c 165 g 212 t  
ORIGIN  
Query Match 20.9%; Score 375.6; DB 12; Length 787;  
Best Local Similarity 67.8%; Pred. No. 3.3e-103;  
Matches 525; Conservative 0; Mismatches 249; Indels 0; Gaps 0;  
QY 205 TCTCCGCGCATGTTTAAAGCCCAAGCAAGAAATCCACATTAACAGATGATTTG 264  
Db 5 TCACACCCAAAACAGAAACACACACTCTCTCTTCTTCACTTCCAAAGTGAATTA 64  
QY 265 TTCACAGAGCGCGCGGAGCGTTGGAGCGGGGAGGTTCTTGTTCAGCCACGAG 324  
Db 65 GTCCAGAAAGCAGACATGCTTTAGATGCTGTAGAAGTCTTTAACTAAACATGA 124  
QY 325 AAGCTACACCGGTTCTCTAAACCGGTGATCTCTAGTGTTCAGATCGCGGAAATTTGCT 384  
Db 125 CTTGAACACCTTTGCGAAACAGCGGACCCACGAGTCCAGATTTCTGGAAATTTGCT 184  
QY 385 CCGGTGATGAACAGCGCGTCCGCGTATCTTCGGTGTGCGAAACCTCCCGATTC 444  
Db 185 CCGGTACCGGAAATTCAGTCTGCAATCTCTTCGGTCAACCGGAAATATCCAAATGT 244  
QY 445 ATCAAAGAGTGTATGTGCGCAACGAGCTAACCCACTTCACGAGCGGTGACAGTCA 504  
Db 245 GTTCAAGCGCTTTAGGTTGAAACGAGGCTAACCTCTTTTGAACCAACCGCGGACAC 304  
QY 505 CACTTCTTCACGAGCGGTATGTTTCAGCGCGTCAAAATTCGAACACGTTTCAGTAGC 564  
Db 305 CATTCTTCACGCGGACGATGTTGTTTCAGCGGTTCAATTTCAAAATGCGTGGTAGT 364  
QY 565 TAGCTTTCGCGGTTTACTCAGATTAACCGGTTTGTTCAGGAACGTCAATTTGGTTCGACCG 624  
Db 365 TAGCTTTCGCGTTTACTGAAACAGAGAGCTTTGTTCAAGAAAGCTTTGGTCCGCT 424  
QY 625 GTTTTCCCAAGCATCGGTGAGCTTCACGCGCACCGGTATTCGCGACTCATGCTA 684  
Db 425 GTTTTCCCTAAAGCCATTGGTGAATATCATGGTCACTCTGGAATTCAGGCTTATGCTG 484  
QY 685 TTCTACGCCAGAGTGCAGCGGTTATAGTCGACCCGCGCACACCGGCTGTAGCTAAC 744  
Db 485 TTTTACGCTCGTGGCTCTTCGGACTTGTGATCACAGTAAGGAAGTGTGTGCAAC 544  
QY 745 GC CGGTTTGTCTAATTTCAATGCGCGGTTATGCGGATGTCGAGGATGATTTACCTTAC 804  
Db 545 GC CGGTTTGTCTAATTTCAATGCGGATTTACTTGTCTAATGTCGAGGATTTGCGCTTAC 604  
QY 805 CAAGTTCAGATCACTCCCAATGGAGATTTAAACACCGTTGCTCGGTTTCGATTTGATGA 864

Db 605 CATGTAAAGGTAACACCCACCGCGATCTTAAACAGAGGGTCTGATTCGATTCGACGCG 664  
QY 865 CAATTAGATCCACATGATTCGCCACCGAAGTCGACCGGAATCCGGTGAATCTTC 924  
Db 665 CAGCTAAATCCACATGATGATCTACCCAAAGTCGACCGGATTCGGTGAATTT 724  
QY 925 GCTTTAAGCTAGCAGCTGCTTTCAAAGCCTTACCTAAATATCTCCGATTCCTA 978  
Db 725 GCTCTTAGCTAGCAGCTGATTCAGAGCATCTCCAGTACTTCAGATTTCA 778  
RESULT 11  
AL950790 667 bp DNA linear GSS 24-OCT-2002  
LOCUS Arabidopsis thaliana T-DNA flanking sequence GK-331D05-016046,  
GENOMIC survey sequence.  
ACCESSION AL950790.1 GI:24407412  
VERSION Arabidopsis thaliana (thale cress)  
KEYWORDS Arabidopsis thaliana  
SOURCE Arabidopsis thaliana  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eutosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE 1 Strizhov,N., Li,Y., Rosso,M., Viehoever,P., Dekker,K., Saedler,H.  
and Weissshaar,B.  
A pipeline for automated high-throughput generation of FSTs  
(flanking sequence tags) from Arabidopsis thaliana T-DNA  
transformed lines  
Unpublished  
2 Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weissshaar,B.  
A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)  
for flanking sequence tag based reverse genetics  
Unpublished  
3 (bases 1 to 667)  
Rosso,M., Strizhov,N., Li,Y. and Weissshaar,B.  
Direct Submission  
Submitted (21-OCT-2002) Weissshaar B., Max-Planck-Institut fuer  
Zuechtungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany  
This sequence is recovered from the left border of the T-DNA. It  
indicates an insertion close to or within gene Atg10100. The  
sequences are generated at the MPI for Plant Breeding Research in  
the context of the GABI-Kat project. GABI-Kat is part of the German  
Plant Genomics program designated 'GABI'. Information on line  
availability can be found at:  
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.  
FEATURES  
Location/Qualifiers  
1..667  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia 0"  
/db\_xref="taxon:3702"  
/clone="GK-331D05-016046"  
/note="Arabidopsis thaliana T-DNA insertion lines"  
/note="PCR was performed on DNA from Arabidopsis thaliana  
plants (T1) which were transformed with the T-DNA from  
vector pAC161. The lines contain one or more T-DNA  
insertions. The DNA fragment(s) resulting from the PCR  
were directly sequenced to determine the genomic sequence  
flanking the insertion. Sequences displaying significant  
similarity to the A. thaliana nuclear genome sequence were  
processed for submission. T-DNA derived sequences were  
removed"  
BASE COUNT 176 a 145 c 161 g 184 t 1 others  
ORIGIN  
Query Match 19.8%; Score 356.2; DB 29; Length 667;  
Best Local Similarity 71.3%; Pred. No. 2.6e-97;  
Matches 469; Conservative 0; Mismatches 189; Indels 0; Gaps 0;  
QY 509 TCTTCGAGGAGCGGTATGTTTCCGCGTCAAAATTCGAACACGGTTCAGTACG 569

```

Db      5  TATTCGACGGTGACGGTATGGTTTACGCGGTAAAAATCACTAAACGAGACGCGAGTTACT 64
Qy      569  CTTCCGGGTTTACTCAGACTACCGGTTTGTTCAGAACGTCATATGGTCGACGCGGTTT 628
Db      65  COTGCGGGTTTACGGAACCGGAGATTTGGTTTCAAGAGAAACAATCCTGGTTCTCCGATTT 124
Qy      629  TCCCAAAAGCCATCGGTGAGCTTTCACGGCCACACCGGTATTGCCGACTCATGTATTTCT 688
Db      125  TCCCTAAAGCTATAGGTGAGCTACATGTCTCTGGAATCGCAGGATGATGCTATTTT 184
Qy      689  AGCCAGAGCTCAGCCGGTATAGTCGACCCGGCACAGGACCGGTGTAGCTAACCGCG 748
Db      185  AGCCAGCGGTTTATTCGGTTTATTAATAACAAAAACGGAACCGGATTCCTAACCGCG 244
Qy      749  GTTTCGTCTATTTCAATGCCCGGTTTATGGCTATGTCGGAGGATGATTACTTACCAG 808
Db      245  GTTTCGTCTATTTCCAGGACCGGTTTATAGCTATGTCGTAAGATGATCTACTTACCAG 304
Qy      809  TTCAGATCACTCCCAATGAGATTTAAAAACCGTTGGTCGGTTCGATTTTATGAGCAAT 868
Db      305  TTCGTGTCACGTACCAATGCGGATTTAGAGACCATCGGAAGATTCGATTTGACGACAAC 364
Qy      869  TAGAATCCACAATGATTGCCACCGGAAAGTCGACCGGAAATCCGGTGAATCTTTCGCTT 928
Db      365  TAAGCTCCGCAATGATCGCTACCCGAGATTGATCCGGTACGAGAGATTTTGGGT 424
Qy      929  TAAGCTACGACTCGTTTCAAGCGCTTACCTTAAATATCTTCGATTCACCGGACGAA 988
Db      425  TGAGCTACGACGCTGTTAAGAAACCGTATTTGAAATATCTTTAAATTCGCGCGGAAGTG 484
Qy      989  CTAATACCGGACGTCGAGATTCAGTTGATCAGCCACGATGATGATGATGATGATGATG 1048
Db      485  AGAATACCGGACGCTGAGATTCCTCTCCGACGTCGACGATGATGATGATGATGATG 544
Qy      1049  TTACAGAGAACTTCGTCGTCGACCTGACCAAGAGTGGTTTCAAGTCGCGGAGATGA 1108
Db      545  TCACCTGAGAATTCGTTGATTCGGATCAACAAGTTGTGTTTAAAGCTCTCCGATATGT 604
Qy      1109  TCCGCGGTGGTCTCCGTTGGTTTACGACAGAAACGATCGCAAGATTCGGGATTTT 1166
Db      605  TTCTTGGGAATATCCCGGTTAAATACGACGGGAGAAATTTCCCGGTTTGGATGTT 662

```

```

RESULT 12
LOCUS   BF113346
DEFINITION  EST440936 tomato breaker fruit Lycopersicon esculentum cDNA clone
          CLE4489 5' sequence, mRNA sequence.
ACCESSION  BF113346
VERSION    BF113346.1 GI:10943036
KEYWORDS   EST.
SOURCE     Lycopersicon esculentum (tomato)
ORGANISM   Lycopersicon esculentum
REFERENCE  1 (bases 1 to 592)
AUTHORS    Alcala, J., Vrebalov, J., White, R., van der Hoeven, R.S., Holt, I.E.,
            Liang, P., Hansen, T.S., Craven, M.B., Bowman, C.L., Renning, C.M.,
            Nierman, W., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley
            , S.D.
TITLE      Generation of ESTs from tomato fruit tissue, breaker stage
JOURNAL    Unpublished
COMMENT    Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall,
            Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
            Location/Qualifiers
            1..592
            /organism="Lycopersicon esculentum"
            /mol_type="mRNA"

```

```

RESULT 13
LOCUS   BZ022957/c
DEFINITION  oeh32c12.g1 B.oleracea002 Brassica oleracea genomic, genomic survey
          sequence.
ACCESSION  BZ022957
VERSION    BZ022957.1 GI:23582691
KEYWORDS   Brassica oleracea
SOURCE     Brassica oleracea
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

```

/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLE4489"
/tissue_type="Pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/clone_lib="tomato breaker fruit"
/note="Vector: pBluescriptSMCquadap; Site 1: EcoRI;
       Site 2: XhoI; supplier: Boyce Thompson Institute;
       sequencing: The Institute for Genomic Research. Fruit
       were harvested at the breaker stage (first sign of
       lycopene accumulation on the blossom end of fruit). Fruit
       were cut in half and the seeds and locules were discarded
       prior to freezing the pericarp."
BASE COUNT  194 a 106 c 131 g 161 t
ORIGIN

```

```

Query Match 19.1%; Score 344; DB 10; Length 592;
Best Local Similarity 73.8%; Pred. No. 1.3e-93;
Matches 437; Conservative 0; Mismatches 155; Indels 0; Gaps 0;
Qy 893 CGAAAGTCACCCGGAATCCGGTGAACCTCTTCGCTTTAAGCTACGACGTCGTTTCAAAGC 952
Db 1 CAAAGCTCGACCCAGTTTCCGGTGAGCTATTTCCTTAGCTACGATGATTCAGAACG 60
Qy 953 CTTACCTAAATACCTTCGATTCACCGGAGGAACTAAATCACCAGGACGTCGAGATTC 1012
Db 61 CATACCTCAAGTACTTCAGATTTTCAAAAATGGGGAATAATCAATGATTTGAATTC 120
Qy 1013 AGTTGATCAGCCACGATGATGACGATTTCCGGAATTACAGAGAACTTCGTCGTCGTAC 1072
Db 121 CAGTTGAAGACCCCAACAATGATGATGATTTCCGAATTACTGAGAATCTTCGTCGTCATTC 180
Qy 1073 CTGACCAAGCAAGTCGTTTCAAGCTCCCGAGATGATCCGGTGGGTCTCCGGTGGTTT 1132
Db 181 CTGATCAACAAGTCGTTTCAAGATCTCTGAAATGATCCGTCGAGGTTTCCCGGTCGTTT 240
Qy 1133 ACCACAAGAAACAAGTCGCAAGATTCGGGATTTTAGCAAAATACGCCGGAAGATTCATCGA 1192
Db 241 ACACAGAAACAAGTTTCCCGATTTGGTATTCGTAAGTACGCGGAAGATGGGTCTG 300
Qy 1193 ACATTAAGTGAATGATGTCGATTCAGATTCGTTTCCATCTCTGGAACGCTTGGGAAG 1252
Db 301 ATTGAATGGGTGAAGTACCTGATTTCTGTTTCTCCCTCTGGAATGCTTGGGAAG 360
Qy 1253 AGCCAGAAACAGATGAAGTCGTCGATAGGGTCTGTATGATCTCCACGACACTCAATTT 1312
Db 361 AAGCAGAAACAGATGAATTCGTTGATTTGGTTCATGATGACACACGACTCCATTT 420
Qy 1313 TCAACGAGTCTGACGAGAACTCTCAAGAGTGTCTCTGTCTGAAATCCGCTGAAATCTCAAAA 1372
Db 421 TCAATGAATGATGAAGGCTAAAGAGTGTTTTATCCGAAATCCGCTCAATTTGAAA 480
Qy 1373 CCGGTGAATCAACTCCCGCTCGATCATCTCCAACGAAGATCAACAGTCAACCTCGAAG 1432
Db 481 CAGGGAATCAACAGAAATCCATTAATCGAAACCCGGATGACCAAGTGAATTTAGAAG 540
Qy 1433 CAGGAGTGTCAACAGAAACATCTCGCGGTAAAAACCAATTCGCTTACTT 1484
Db 541 CTGGAATGGTGAACCCGAAACAACTCGGAAGGAAACACAGATGATGCTTATTT 592

```

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; eurosids II; Brassicales; Brassicaceae; Brassica.

## REFERENCE

1 (bases 1 to 696)  
Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T., Nash  
W., Rabinowicz,P.D. and Wilson,R.K.  
Whole genome shotgun reads from Brassica oleracea

## TITLE

Unpublished

## JOURNAL

Contact: Richard K. Wilson

## COMMENT

Genome Sequencing Center

Washington University School of Medicine

Email: submissions@watson.wustl.edu

Plate: oeh32 row: c column: 12

Seq primer: -28RPpOT reverse

Class: shotgun

High quality sequence start: 16

High quality sequence stop: 551.

## FEATURES

source

1..696

/organism="Brassica oleracea"

/mol\_type="genomic DNA"

/db\_xref="taxon:3712"

/clone\_lib="B.oleracea002"

/note="Vector: pOTw13; Whole genome shotgun library from

flowering buds. DNA was purified from a crude nuclear

prep using Brassica oleracea 101000DH3 buds provided by

Thomas Osborn at the University of Wisconsin. genomic

DNA was provided by Pablo Rabinowicz (CSHL) and the

shotgun library prepared at Washington University Genome

Sequencing Center."

150 a 205 c 161 g 180 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 18.6%; Score 334.2; DB 28; Length 696;

Mismatches 441; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY

793 GATTTCCTTACCAAGTTCAGATCACTCCCAATGGAGATTAAACCGTTGTCGGTTC 852

Db

696 GATTTCGGTATCAGCTGAATAAATCAACACCGGGAATCTCGAACCGTTGGCGTTAC 637

QY

853 GATTTCGATGGACAATTAGAAATCAACAATGATTCGCCACCGGAATGACCGCGGAATCC 912

Db

636 GATTTCACGGTCACTTAAATATCACGATATCGCCACCGCAATCTCGACCGGTATCG 577

QY

913 GGTGAATCTTCGTTTAAGCTACGAGCTGTTTCAAGCTTACCTTAAATCTTCCGA 972

Db

576 AAGGAGCTCCACGGCTTAAGCTACGAGCTGTCGAGAAAGCTTACTTGAAGTACTTCAGA 517

QY

973 TTCTCACCAGCGGAATCAATACACCGGACGTCGAGATTGAGTTGATCAGCAAGATG 1032

Db

516 TTCTCGCGGAGCGGTTAAATACCCGGAATCTAGAGATCCCGTTCGAGACTCCGAGATG 457

QY

1033 ATGCAGATTCGGGATTACAGAGAACTTCGTGCTGCTACTGACGAGCAAGTCGTTTC 1092

Db

456 ATTCACGATTCGTTATACGAGAACTTCGTGTTGATTCGGATCAACAGTCGTGTTTC 397

QY

1093 AAGCTGCGGAGATGATCCGGGTGGGTCTCCGGTGTGTTTACGACAGAAACAAGTCGCA 1152

Db

396 AAGTAGGGAGATGATGCGCGGAACTCGCGGTGTTTTCGACGCGGATAGGTTTCG 337

QY

1153 AGATTGGGATTTAGACAAATACCGGAGATTCATCGAACATTAAGTCGATTGATGCT 1212

Db

336 CGATTGGGGATAATGCCGAAAGACGCGACGGAGCTTCGAGATAATCTCGGTGGATTTCG 277

QY

1213 CCAGATTCTCTGCTCCATCTCTGGAACCGTTGGGAAGACGCGAGAAACAGATGAAGTC 1272

Db

276 CCGGAGAGCTTCTGTTTCATCTCTGNAACGCTGGGNAATCGCCGAGACTGAGGAATG 217

QY

1273 GTCGTGATAGGTCCTGATGATCTCCACAGACTCAATTTTCAACAGTCTGACGAGAAAT 1332

Db

216 TTGTTGATCGGTCTGATGTCGCGCGGATTTCTATCTTCAACGAGAGACGAGAGC,157

QY

1333 CTCAGAGTGTCTCTGAAATCGCGCTGAATCTCAAAACCGGTGAATCAACTCGCCGT 1392

156 TTGAAAAGCGTCTTTGACGGAGATCAGATAAACCTCAGGACGCTGAATCCACGCGCGA 97

QY

1393 CCATCATCTCCAAACGAAG 1411

Db

96 GCGATGATGTCGACGATG 78

## RESULT 14

BE461924

LOCUS

DEFINITION

ESR143439 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA

clone cLEG40L2, mRNA sequence.

ACCESSION

BE461924

VERSION

BE461924.1

KEYWORDS

EST.

SOURCE

ORGANISM

Lycopersicon esculentum (tomato)

Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.

1 (bases 1 to 566)

Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,

Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Roming,C.M.,

Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley

,S.D.

Generation of ESTs from tomato fruit tissue, breaker stage

Unpublished

Contact: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: http://www.genome.clemson.edu/orders/index.html

5 prime sequence.

Location/Qualifiers

1..566

/organism="Lycopersicon esculentum"

/mol\_type="mRNA"

/cultivar="TA496"

/db\_xref="taxon:4081"

/clone="cLEG40L2"

/tissue\_type="Pericarp"

/dev\_stage="breaker"

/lab\_host="SOLR"

/clone\_lib="tomato breaker fruit, TIGR"

/notes="Vector: pBluescriptSKM2adapt; Site.1: EcoRI;

Site.2: XhoI; Fruit were harvested at the breaker stage

(first sign of lycopene accumulation on the blossom and

the fruit). Fruit were cut in half and the seeds and

locules were discarded prior to freezing the pericarp."

184 a 104 c 124 g 154 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 18.5%; Score 332.4; DB 10; Length 566;

Mismatches 420; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY

893 CGAAAGTCGACCGGAATCCGGTGAACTCTTCGCTTTAAGCTAGACGTCGTTCAAAGC 952

Db

1 CAAGCTCGACCCAGTTTCCGGTGAGCTATTTCCTTAGCTACGATGTGATTGAGAAGC 60

QY

953 CTTACTTAATATCTTCGATTCCTCACCGGACGAACTAAATCACCGACGTCGAGATTC 1012

Db

61 CATACCTCAAGTACTTCAGATTTTCAAAATGGGAAATCAATGATGTTGAAATTC 120

QY

1013 AGCTTGATCAGCAACGATGATGACGATTTGGGATTACAGAACTTCGTCGTCGTAC 1072

Db

121 CAGTTGGAAGACCAACAATGATGATGATTTCCGCAATTTACTGAGAACTTCGTCGTCATTC 180

QY

1073 CTGACACGACGATCGCTTTTCAAGCTCCGAGATGATCCGCGGTGGGTCTCCGGTGGTTT 1132

Db

181 CTGATCAACAGTCGCTTTTCAAGATGTCGAAATGATCCGTGGAGGTTCAACGCGTGGTTT 240

```

QY 1133 ACGACAGACAGAGTCCGAGATTCGGGATTTAGACAAATACGCCGAGATTCATCGA 1192
Db 241 ACGACAGACAGAGTTCGCCGATTTGGTATTTGGATAAGTAGCGGAAAGATGGGTCTG 300
QY 1193 ACATTAAAGTGATGATGCTCCAGATTCCTTCTGCTTCCATCTCTGGAACGCTTGGGAAG 1252
Db 301 ATTTGAAATGGTTGAAGTACCTGATTTCTGTTTCCACCTCTGGAATGCTTGGGAAG 360
QY 1253 AGCCAGAACAGATGAAGTCTGCTGATAGGCTCTGTATGACTCCACGAGTCAATTT 1312
Db 361 AAGCAGAACAGATGAATCTGTTAATGTTGTTTATGATGACACACGAGTCCATTT 420
QY 1313 TCACAGAGTCTGACGAGAAATCTCAAGAGTGTCTCTGTAATCCGCTCGAATCTCAAAA 1372
Db 421 TCAATGATGATGATGAAGGCTAAGAGTGTGTTTATCCGAAATCCGCTCAATTTGAAA 480
QY 1373 CCGGTGAATCAATCGCGTCCGATCATCTCCACGAGAAATCAACAGTCAACTCGAAG 1432
Db 481 CAGGAAATCAACAGAAATCCATATCGAAACCCGGATGAACAAAGTGAATTTGAAG 540
QY 1433 CAGGATGCTCAACAGAAACATGCTC 1458
Db 541 CTGGAATGCTCAACGAAACAACTC 566

RESULT 15
BU550566/c
LOCUS
DEFINITION GM880021A20F06 Gm-r1088 Glycine max cDNA clone Gm-r1088-7932 3',
mRNA sequence.
ACCESSION BU550566
VERSION BU550566.1 GI:22933427
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; euroside 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE
1 (bases 1 to 637)
Vodkin,L., Shoemaker,R., Keim,P., Retzel,E., Khanna,A., Shealy,R.,
Clough,S., Thibaud-Nissen,P., Coryell,V., Erpelding,J., Raph,C.,
Shoop,E., Stromvik,M., Schweitzer,P., Gong,G. and Liu,L.
A Functional Genomics Program for Soybean (NSF 9872565) (2002)
Unpublished
Other ESTs: BR99117 corresponding to Gm-cl068-3283 (5')
Contact: Vodkin, L.O., PI, A Functional Genomics Program for
Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional
Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
Insert Length: 637 Std Error: 0.00
Plate: GM880021A20 row: F column: 06
Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'
High quality sequence stop: 637.
location/Qualifiers
1..637
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="Gm-r1088-7932"
/clone_lib="Gm-r1088"
/clone="The library Gm-r1088 is a sequence-driven, reracked
set of 9,216 low redundancy clones selected from cDNA
libraries from various tissues and stages of development
of soybean that consists of 2,706 cDNAs from germinating
cotyledons (source library Gm-cl027); 1,355 cDNAs from
immature seed coats (libraries Gm-cl019 and Gm-cl023); 917
cDNAs from tissue culture derived somatic embryos (source

```

libraries Gm-cl036 and Gm-cl075); 3,275 cDNAs from germinating seedlings, shoot tips, or leaves exposed to various stresses (source libraries Gm-cl064, Gm-cl065, Gm-cl066, and Gm-cl067; and Gm-cl068); and 963 cDNAs from young leaves exposed to bacterial and fungal pathogens (source libraries Gm-cl072, Gm-cl073, and Gm-cl074). The 5' ESTs of the source clones from the different libraries was used to select singletons, or a representative of each contig, which were rackted to form library Gm-r1088 and the cDNA clones of the rackted Gm-r1088 library were then sequenced at the 3' end. The unigene selection and 3' sequencing was funded by NSF Plant Genome project #9872565 (http://soybean.genomics.croplsci.uiuc.edu/) as part of creation of a low redundancy soybean cDNA set. The source cDNA libraries were constructed by the laboratories of Lila Vodkin, University of Illinois, Randy Shoemaker, Iowa State University, and Paul Keim, Northern Arizona University as part of the Public EST project, http://129.186.26.94/soybeanest.html. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, http://web.ahc.umn.edu/biodata/nsfsoy/. Reracking and 3' sequencing were conducted by services of the University of Illinois Keck Center for Comparative and Functional Genomics http://www.biotech.uiuc.edu/keck.htm. Note: The corresponding 5' EST from each clone in the Gm-r1088 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'.

```

BASE COUNT 160 a 180 c 126 g 171 t
ORIGIN
Query Match 18.1%; Score 326; DB 13; Length 637;
Best Local Similarity 72.1%; Pred. No. 4.5e-88;
Matches 454; Conservative 0; Mismatches 170; Indels 6; Gaps 2;
QY 1140 GAACAGGTGCGCAAGATTCGGGATTTAGACAAATACGCCGAGATTCATCGAACTTAA 1199
Db 637 GAACAGGTTCACAGATTCGGGATTCGACAAAGATCTAAAGATGCAAAATGATATGAA 578
QY 1200 GTGGATTGATGCTCCAGATTCGTTCTGTTCCATCTCTGGAACGCTTGGGAAGACCCAGA 1259
Db 577 ATGGATCGATGCTCCAGATTCGTTCTGTTCCATCTCTGGAACGCTTGGGAAGACCCAGA 518
QY 1260 AACAGATGAAGTCGTGCTGATAGGTCCTGTATGACTCCACAGACTCAATTTTCAACA 1319
Db 517 AATGATGAGATCGTTGTTCATCGGTTCTCTGATGACCCCGCGGACTCCATTTTCAACGA 458
QY 1320 GTCTGACGAGATCTCAAGATGTCTGTCTGAAATCCGCTGAAATCTCAAAACCGGTGA 1379
Db 457 ATCGGAAGAGATTTGAAGAGCATTTGCGAGATAGGCTGAATTTGAAGACAGGCA 398
QY 1380 ATCAACTCCCGCTCCGATCATCTCCAAACGAAAGATCAACAAAGTCAACCTCGAAGCAGGAT 1439
Db 397 GTCCACGAGAAAACCCATTTATCTC---GGAATCCGAAACAAAGTCAACTTGGAAAGCCGGAT 341
QY 1440 GGTCAACAGAACATGCTCGGCCGTAAACCAATTCGTTACTTGGCTTTTACCGAGCC 1499
Db 340 GGTCAACAGAAAACAGCTCGGGAGAAAGACGAGTTCCGGGTACTTAGCACTTTCGGGAGCC 281
QY 1500 GTGGCCCTAAAGTCTCAGGATTCGCTAAAGTTGATCTCACTACTGGAGAAAGTTAAGAAACA 1559
Db 280 GTGGCCCTAAGTTTCGGGTTTTCGCAAGTTTGTATTTGTTTCAGTGGGAGTTTAAAGAAGTA 221
QY 1560 TCTTTACGGGATTAACCGTTAGCGAGAGAGCTCTGTTTCTCCC---CGGAGAAAGAGG 1616
Db 220 CATGTATGCGAAGAGAGGTTTCGGAGGGAGGCTCTGTTTCTTCCAAATGGTGTGCAAGG 161
QY 1617 AGAGGAAGACGAGAGGATACATCTCTGTTTCTGTTTCCACGACGAGACATCGAAATTCGGA 1676

```



Db 160 TGATGAAGACGATGGCCATATTCTCGCATTCGTCATGATGAGAAAGAGTGCATAATCGGA 101  
Qy 1677 GTTACAGATAGTTAACCGGTTAGCTTAGAGTTGAGCAACGGTTAAACTTCGGTCAAG 1736  
Db 100 GCTGCAGATTGTCATGCCCAAACTTTGAAGCTTGAGGCTTCAGTTAGCTTCCTTCCAG 41  
Qy 1737 GGTTCGGTACGGATTTTCACGGTACATTCAT 1766  
Db 40 AGTTCCTTATGGGTTTCACGGTACTTTAT 11

Search completed: November 16, 2003, 18:30:16  
Job time : 3974 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2003, 18:30:22 ; Search time 593 Seconds  
(without alignments)

9920.351 Million cell updates/sec

Title: US-09-758-269-5

Perfect score: 1800  
Sequence: 1 atgcttcttttcacggcaac.....tggcgaagcaggctgtgtga 1800

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

```
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1800	100.0	1800	10	US-09-758-269-5
2	824.6	45.8	1818	10	US-09-758-269-15
3	798.4	44.4	1839	10	US-09-758-269-11
4	738.2	41.0	1752	10	US-09-758-269-1
5	606.6	33.7	1815	10	US-09-758-269-13
6	528.2	29.3	1734	10	US-09-758-269-9
7	138	11.0	200	9	US-09-770-698-172
8	190.4	10.6	393	10	US-09-878-574-2543
9	169.6	9.4	1788	10	US-09-758-269-3
10	169.6	9.4	1788	10	US-09-938-842A-1444
11	139.8	7.8	1617	10	US-09-758-269-7
12	139.8	7.8	1617	10	US-09-758-269-17
13	93	5.2	320	10	US-09-878-574-2872
14	60.8	3.4	367	9	US-09-770-791-402
15	50	2.8	1518	8	US-08-976-063C-21
16	50	2.8	32679	8	US-08-976-063C-1

```

17 40.4 2.2 534 12 US-10-149-759-65 Sequence 65, Appl
18 38.8 2.0 1416 9 US-09-815-242-4160 Sequence 4160, Ap
19 36.2 2.0 699 9 US-09-815-242-4096 Sequence 4096, Ap
20 35.6 2.0 594 12 US-10-140-472-10 Sequence 10, Appl
21 35.6 2.0 594 12 US-10-141-761-10 Sequence 10, Appl
22 35.6 2.0 594 12 US-10-142-885-10 Sequence 10, Appl
23 35.6 2.0 594 12 US-10-158-790-10 Sequence 10, Appl
24 35.6 2.0 594 12 US-10-137-871-10 Sequence 10, Appl
25 35.6 2.0 594 12 US-10-140-805-10 Sequence 10, Appl
26 35.6 2.0 594 12 US-10-140-864-10 Sequence 10, Appl
27 35.6 2.0 594 12 US-10-140-923-10 Sequence 10, Appl
28 35.6 2.0 594 12 US-10-141-756-10 Sequence 10, Appl
29 35.6 2.0 594 12 US-10-141-759-10 Sequence 10, Appl
30 35.6 2.0 594 14 US-10-123-155-10 Sequence 10, Appl
31 35.6 2.0 594 15 US-10-146-731-10 Sequence 10, Appl
32 35.6 2.0 882 14 US-10-184-644-574 Sequence 574, App
33 35.6 2.0 882 14 US-10-184-634-574 Sequence 574, App
34 35.6 2.0 1722 12 US-10-316-253-208 Sequence 208, App
35 35.6 2.0 2140405 12 US-10-027-632-76212 Sequence 76212, A
36 35.6 2.0 2140405 13 US-10-027-632-76212 Sequence 76212, A
37 35.4 2.0 671 14 US-10-184-644-346 Sequence 346, App
38 35.4 2.0 671 14 US-10-184-634-346 Sequence 346, App
39 35 1.9 409 11 US-09-918-935-33359 Sequence 33359, A
40 35 1.9 598 12 US-10-027-632-208895 Sequence 208895,
41 35 1.9 598 12 US-10-027-632-208896 Sequence 208896,
42 35 1.9 598 13 US-10-027-632-208895 Sequence 208895,
43 35 1.9 598 13 US-10-027-632-208896 Sequence 208896,
44 34.8 1.9 4695 10 US-09-969-708-37 Sequence 37, Appl
45 34.6 1.9 572 10 US-09-878-178-1534 Sequence 1534, Ap

```

#### ALIGNMENTS

#### RESULT 1

```

US-09-758-269-5
; Sequence 5, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT FILING DATE: 2001-01-12
; PRIOR FILING DATE: 2001-01-11
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR APPLICATION NUMBER: JP 2000-010056
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1800
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1797)
US-09-758-269-5

```

Query Match 100.0%; Score 1800; DB 10; Length 1800;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1800; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCTTCTTTTCACGGCAACGGCTGGGTTCTTGGAGATGGCTTGGTGGCAATCATCTACT 60

Db 1 ATGCTTCTTTTCACGGCAACGGCTGGGTTCTTGGAGATGGCTTGGTGGCAATCATCTACT 60

Qy 61 CAGCCGCATTATTCGTTCTTCTCAAGTCGACTTATTTAGCTTCTTACCTATG 120

Db 61 CAGCCGCATTATTCGTTCTTCTCAAGTCGACTTATTTAGCTTCTTACCTATG 120

1201 TGGATTGATGCTCCAGATTGCTTCTGCTTCCATCTCTGAAACGCTTGGAGAGCCAGAA 1260  
 Db  
 1201 TGGATTGATGCTCCAGATTGCTTCTGCTTCCATCTCTGAAACGCTTGGAGAGCCAGAA 1260  
 Qy  
 1261 ACAGATGAAGTCTGCTGATAGGCTTCTGATGATGCTCCACAGAGCTCAATTTTCAACGAG 1320  
 Db  
 1261 ACAGATGAAGTCTGCTGATAGGCTTCTGATGATGCTCCACAGAGCTCAATTTTCAACGAG 1320  
 Qy  
 1321 TCTGACGAGAATCTCAAGAGTGTCTCTGCTGCTGAAATCCGCTGAAATCTCAAAACCCGCTGAA 1380  
 Db  
 1321 TCTGACGAGAATCTCAAGAGTGTCTCTGCTGCTGAAATCCGCTGAAATCTCAAAACCCGCTGAA 1380  
 Qy  
 1381 TCAACTCGCGCTCCGATCATCTCCACGAGAGTCAACCTCGAAGCAGGAGTG 1440  
 Db  
 1381 TCAACTCGCGCTCCGATCATCTCCACGAGAGTCAACCTCGAAGCAGGAGTG 1440  
 Qy  
 1441 GTCAACAGAAACATGTCTCGCGCTGAAACCAAAATTCGCTTACTTGGCTTTAGCCGAGCGG 1500  
 Db  
 1441 GTCAACAGAAACATGTCTCGCGCTGAAACCAAAATTCGCTTACTTGGCTTTAGCCGAGCGG 1500  
 Qy  
 1501 TGGCTTAAAGTCTCAGGATTCGCTTAAAGTTGATCTCACTACTGGAGAGTTAAGAAACAT 1560  
 Db  
 1501 TGGCTTAAAGTCTCAGGATTCGCTTAAAGTTGATCTCACTACTGGAGAGTTAAGAAACAT 1560  
 Qy  
 1561 CTTTACGCGCATAAACCGTTACGAGGAGAGCCCTCTCTTTCTCCCGGAGAGGAGGAG 1620  
 Db  
 1561 CTTTACGCGCATAAACCGTTACGAGGAGAGCCCTCTCTTTCTCCCGGAGAGGAGGAG 1620  
 Qy  
 1621 GAAGACGAGATACATCTCTGCTTTCGTTTACGAGAGAGATCGGAATCGGAGTTA 1680  
 Db  
 1621 GAAGACGAGATACATCTCTGCTTTCGTTTACGAGAGAGATCGGAATCGGAGTTA 1680  
 Qy  
 1681 CAGATAGTTAACGCGCTTACGTTAGAGTTGAAGCAACGGTTAAACTTCCGTCAGGGTT 1740  
 Db  
 1681 CAGATAGTTAACGCGCTTACGTTAGAGTTGAAGCAACGGTTAAACTTCCGTCAGGGTT 1740  
 Qy  
 1741 CCGTACGAGTTTACGCTACATTCATCGGAGCCGATGATTTGGCGAGCAGGTCGTGTGA 1800  
 Db  
 1741 CCGTACGAGTTTACGCTACATTCATCGGAGCCGATGATTTGGCGAGCAGGTCGTGTGA 1800

RESULT 2

US-09-758-269-15  
 ; Sequence 15, Application US/09758269  
 ; Patent No. US20020104120A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: IUCHI, SATOSHI  
 ; APPLICANT: KOBAYASHI, MASATOMO  
 ; APPLICANT: SHINOZAKI, KAZUO  
 ; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN  
 ; FILE REFERENCE: 3914-3  
 ; CURRENT APPLICATION NUMBER: US/09/758,269  
 ; PRIOR FILING DATE: 2001-01-12  
 ; PRIOR APPLICATION NUMBER: JP 2001-003476  
 ; PRIOR FILING DATE: 2001-01-11  
 ; PRIOR APPLICATION NUMBER: JP 2000-010056  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 15  
 ; LENGTH: 1818  
 ; TYPE: DNA  
 ; ORGANISM: Lycopersicon esculentum  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(1815)  
 US-09-758-269-15

Query Match 45.8%; Score 824.6; DB 10; Length 1818;  
 Best Local Similarity 69.6%; Pred. No. 1.3e-273;  
 Matches 1134; Conservative 0; Mismatches 489; Indels 6; Gaps 1;

121 GCCAGTGTGTACACGTAAGTCAATGTTTTCATCTGCGCTTCACTCTCTCGAGCTTT 180  
 Db  
 121 GCCAGTGTGTACACGTAAGTCAATGTTTTCATCTGCGCTTCACTCTCTCGAGCTTT 180  
 Qy  
 181 CATTTCCCTTAAGCAATCATCAAACTCTCCCGCAATGTTTAAAGCCCAAGCAAGAA 240  
 Db  
 181 CATTTCCCTTAAGCAATCATCAAACTCTCCCGCAATGTTTAAAGCCCAAGCAAGAA 240  
 Qy  
 241 TCAACACTAACAAGATGATGTTTTCAGAGAGCGCGGCGAGCGTTGAGCGCGCG 300  
 Db  
 241 TCAACACTAACAAGATGATGTTTTCAGAGAGCGCGGCGAGCGTTGAGCGCGCG 300  
 Qy  
 301 GAGGTTTCTCTGTGAGCCACAGAGTCAACCCGCTTCTTAAACCGCTGATCCTAGT 360  
 Db  
 301 GAGGTTTCTCTGTGAGCCACAGAGTCAACCCGCTTCTTAAACCGCTGATCCTAGT 360  
 Qy  
 361 GTTCAGATCGCGGAAATTTTGTCTCGGTGAATGAACACCCGCTCGCGGTAAATCTTCG 420  
 Db  
 361 GTTCAGATCGCGGAAATTTTGTCTCGGTGAATGAACACCCGCTCGCGGTAAATCTTCG 420  
 Qy  
 421 GTGGTCGGAACCTTCCCGATTCATCAAGAGGTGATGTGCGCAACGAGCTAAACCA 480  
 Db  
 421 GTGGTCGGAACCTTCCCGATTCATCAAGAGGTGATGTGCGCAACGAGCTAAACCA 480  
 Qy  
 481 CTTTACGAGCCGCTGACAGTCAACCTCTTTCGAGCGAGAGCGGTATGTTTCAACCGT 540  
 Db  
 481 CTTTACGAGCCGCTGACAGTCAACCTCTTTCGAGCGAGAGCGGTATGTTTCAACCGT 540  
 Qy  
 541 AAATTCGAAACAGCGTTTCACTAGCTAGCTTCCCGGTTTACTCAGACTAACCGGTTTGT 600  
 Db  
 541 AAATTCGAAACAGCGTTTCACTAGCTAGCTTCCCGGTTTACTCAGACTAACCGGTTTGT 600  
 Qy  
 601 CAGGAACGTCATTTGGTTCGACCGGTTTCCCAAGCCATCGGTGAGTTTCAACCGCAC 660  
 Db  
 601 CAGGAACGTCATTTGGTTCGACCGGTTTCCCAAGCCATCGGTGAGTTTCAACCGCAC 660  
 Qy  
 661 ACCGTTATCCCGACTGATGCTATTTCTAGCGCAGAGCTGACCGGATAGTCGACCGG 720  
 Db  
 661 ACCGTTATCCCGACTGATGCTATTTCTAGCGCAGAGCTGACCGGATAGTCGACCGG 720  
 Qy  
 721 GCACACGGAACCGGTGTAGTAAACCGGTTTGGTCTTATTCATGCGCGGTTATTTGGT 780  
 Db  
 721 GCACACGGAACCGGTGTAGTAAACCGGTTTGGTCTTATTCATGCGCGGTTATTTGGT 780  
 Qy  
 781 ATGTGCGAGGATGATTTACCTTCAACAGTTTCACTTCCCAATGAGATTTAAACCC 840  
 Db  
 781 ATGTGCGAGGATGATTTACCTTCAACAGTTTCACTTCCCAATGAGATTTAAACCC 840  
 Qy  
 841 GTTGTGCGGTTTGGATTTGATGGACAATTAGAAATCCCAATGATTTGCCACCCGAAAGTC 900  
 Db  
 841 GTTGTGCGGTTTGGATTTGATGGACAATTAGAAATCCCAATGATTTGCCACCCGAAAGTC 900  
 Qy  
 901 GACCCGGAATCCGCTGAACTTCTGCTTTAAGCTACGAGTCTGTTTCAAGCCCTTACCTA 960  
 Db  
 901 GACCCGGAATCCGCTGAACTTCTGCTTTAAGCTACGAGTCTGTTTCAAGCCCTTACCTA 960  
 Qy  
 961 AAATACCTCCGATTTCAACCGGAGCACTAAATCAGCGAGCTGAGATTCAGCTTGTAT 1020  
 Db  
 961 AAATACCTCCGATTTCAACCGGAGCACTAAATCAGCGAGCTGAGATTCAGCTTGTAT 1020  
 Qy  
 1021 CAGCCACGATGATGACGAGATTTCCGATTTACAGAGAACTTCTGCTGCTGCTGACCTGACCAG 1080  
 Db  
 1021 CAGCCACGATGATGACGAGATTTCCGATTTACAGAGAACTTCTGCTGCTGCTGACCTGACCAG 1080  
 Qy  
 1081 CAGCTCGTTTTCAGCTGCGGAGATGATCCCGGTTGCTCCGCTGGTTTACGACAAG 1140  
 Db  
 1081 CAGCTCGTTTTCAGCTGCGGAGATGATCCCGGTTGCTCCGCTGGTTTACGACAAG 1140  
 Qy  
 1141 AACAGAGTCGAAGATTCGGATTTTACACAAATACGCGGAGATTCATCGAATTAAG 1200  
 Db  
 1141 AACAGAGTCGAAGATTCGGATTTTACACAAATACGCGGAGATTCATCGAATTAAG 1200

Qy	170	CTCCAGCTCTTCATTTCCCTAAGCAATCATCAAACTCTCCCGCCATTTGTTTAAAGCCCA	229
Db	185	CTTCAAATTTATCAAAACACCAACAAAGAAATAATACAAATTTACACCCAAAACAAAGAAACAACA	244
Qy	230	AAGCCAAAGAAATCCACACACTAACAGATGAATTTGTTCCAGAGAGCGCGCGGACAGGT	289
Db	245	ACTCTCTTCTTCTTCCAATCTTCAAGTGGAAATTTAGTGCAGAAACAGCAGCAATGGT	304
Qy	290	TGGACGCGCGGAGGGTTCCTTGTTCAGCCACGAGAAGCTACACCCGGTTCTCTAAAAACGG	349
Db	305	TAGATGCTGTAGAAAGTGCTTTAACTTAAACATGAACCTTGAACACCGCTTTCGGAACACG	364
Qy	350	CTGATCCTAGTGTTCAGATCGCGGAAATTTTGTCTCCGGTGAATGAACAGCCCGTCCGGC	409
Db	365	CGACCCACAGTCCAGATTTCTGGGAATTTTGTCTCCGGTACCGGAAATTCACGTCGTGC	424
Qy	410	GTAATCTTCGCGTGTTCGAAACCTTCCGGAATTCATCAAGAGAGTGTATGTCGCAACG	469
Db	425	AATCTCTTCGCGTACCGGAAATAATCCAAATGTGTTCAAGCGTTTACGTTCCGAAACG	484
Qy	470	GAGCTAACCCACTTTCAGAGCGGGTACAGGTCAACACTTCTTCGACGAGAGCGTATGG	529
Db	485	GAGCTAACCCCTTTTGTGAACCAACCGCGGACACCATTTCTTCGACGCGAGCGGTATGG	544
Qy	530	TTCAACGCGCTCAAAATTCGAACACGGTTCAGCTAGCTACGTTTCCGGTTTACTTCAGACTA	589
Db	545	TTCAACGCGTTCAAATTCMAAATGGGTTCGGTAGTTACGTTTTCGCTTCACTGAAACAG	604
Qy	590	ACCGGTTTGTTCAGGAACGTCAATTTGGTTCGACCGGTTTTCCCAAGCCATTCGTGAGC	649
Db	605	AGAGGCTTGTTCAGAAAAAAGCTTTGGTTCGCCCTGTTTTCCTTAAAGCCATTTGGTGAAT	664
Qy	650	TTCAAGGCCACACCGGTATTTGCCGACTCATGTATTTACGCCAGAGCTGCAGCCGGTA	709
Db	665	TACATGTCCTCTGGAAATTCGAAGCTTATGCTGTTTACGCTCGTGGGCTCTTCGGAC	724
Qy	710	TAGTCAGCCCGGCACACGGAAACCGGTGTAGCTAACCGCGGTTTGGTCTATTTCAATGGCC	769
Db	725	TTGTGATCACAGTAAGAGAACTGGTGTTCAAAACGCGGTTTAGTCTATTTCAATTAACC	784
Qy	770	GGTATTTGGCTATGTCGGAGAGATTTACTTACCAAGTTTCAGATCACTCCCAATGGAG	829
Db	785	GATTACTTGCTATGCTGAAGATGATTTGCTTTACCATGTAAAGGTAAACACCACCGCGG	844
Qy	830	ATTTTAAAAACGGTTTGGTCGTTTCGATTTTGATGGACAATTTAGAAATCACAAATGATTCGCC	889
Db	845	ATCTTAAAAACAGAGGGTCGATTCGATTTACAGCGCCAGCTTAAAAATCACCATGATAGCTC	904
Qy	890	ACCGAAAGTCGACCCGGAATCCGGTGAATCTCTTCGCTTTTAAAGCTAGCGGTGTTCAA	949
Db	905	ACCCAAAGCTCGACCCAGTTTCCCGGTGAGCTATTTGCTCTTAGCTAGATGTGATTCAGA	964
Qy	950	AGCCTTACCTAAAAATCTTCCGATTTCTCACCGGACGGAACCTAAATCACCGGAGCTCGAGA	1009
Db	965	AGCCATACCTCAAGTACTTCAGATTTTCAAAAAATGGGAAAAATCAATGATGTTGAAA	1024
Qy	1010	TTGAGCTTGATCAGCAACGATGATGACGATTTTCGGATTTACAGAGAACTTCGTGGTCG	1069
Db	1025	TTCCAGTTGAAGACCCCAACATGATGATTTTCGAAATTTACTGAGAACTTCGTGCTCA	1084
Qy	1070	TACCTGACCAAGTCCGTTTTCGAGTCCCGAGAGATGATCCGCGGTGGGTCTCCGCTGG	1129
Db	1085	TTCTGATCAACAGTCCGTTTTCAGATGCTCGAAATGATCCGTTGAGGTTTACCCTGG	1144
Qy	1130	TTTACGACAAGAACAGGTCCGAGATTCGGGATTTTACAGAAATACGCCGAGATTCAT	1189
Db	1145	TTTACGACAAGAACAAAGTTTCCGATTTGGTATTTCTGGATAAGTACGCCGAAAGATGGGT	1204
Qy	1190	CGAACATTTAAGTGGATTCATGCTCCAGATTCGCTTCTGCTTCCATCTCTGGAACCTTGGG	1249
Db	1205	CTGATTTTGAATGGGTTGAAGTACCTGATTTGTTCTGTTTCCACTCTGGAATGCTTGG	1264
Qy	1250	AAGAGCCGAAACAGATGAAGTCGCTGTAAGGGTCTGTATGACTCCACAGACTCAA	1309

DEPT. T. 3

```

US-09-758-269-11
Sequence 11, Application US/09758269
Patent No. US2002010420A1
GENERAL INFORMATION:
APPLICANT: IUCHI, SATOSHI
APPLICANT: KOBAYASHI, MASATOMO
APPLICANT: SHINOZAKI, KAZUO
TITLE OF INVENTION: TRANSGENIC PLANTS CAPABLE OF PRODUCING A
TITLE OF INVENTION: CLEAVAGE ENZYME GENE
FILE REFERENCE: 3914-3
CURRENT APPLICATION NUMBER: US/09758,269
CURRENT FILING DATE: 2001-01-12
PRIOR APPLICATION NUMBER: JP 2001-003476
PRIOR FILING DATE: 2001-01-11
PRIOR APPLICATION NUMBER: JP 2000-010056
PRIOR FILING DATE: 2000-01-13
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 11
LENGTH: 1839
TYPE: DNA
ORGANISM: Vigna unguiculata
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1836)
US-09-758-269-11

```

Query March 44.4%; Score 798.4; DB 10; Length 1839;  
 Best Local Similarity 70.4%; Pred. No. 1.5e-254;  
 Matches 1098; Conservative 0; Mismatches 456; Indels 6; Gaps 2;  
 QY 233 CCAAGAGATCCAACTTAAACAGATGATTTGTCACAGAGCGCGCGGACGGCTGG 292

Db 281 CCRACCAACATTAACCTCAAAATGGAATCTTCTCAGAAAGCGCTGCCAGCGCTTGG 340  
Qy 293 ACGCGCGAGGTTTCTTGTGACGACGAGAGCTACACCGCTTCTTAAACGCGTG 352  
Db 341 ACCTGGTCAAAACGCGCTCTGTCTGACGAGCGCAACACCGCTCCCAAAACGCGG 400  
Qy 353 ATCTAGTGTTCAGATCGCGGAAATTTTCTCGGTGAATGAACAGCCCGTCCGCGTA 412  
Db 401 ACCGAGGTCCTCAATCGCCGGAATCTTCGCGCGGTGCGGAGCATGCGCGGATCAAG 460  
Qy 413 ATCTTCGCTGTCGAAATCTTCCGATTCATCAAAAGAGTGTATGTGCGCAACGAG 472  
Db 461 GACTCCCGGTGTCGAAATATCCCAATTCATTCACGCGGTGTACGTGCGCAACGSGT 520  
Qy 473 CTAACCCACTTCACGAGCGGTGACAGGTCAACCACTTCTTCGACGAGAGCGGTATGGTTC 532  
Db 521 CCAATCCGCTCTACGAGCGTGTGCGCGGCAACCACTTCTTCGACGCGAGCGCATGGTCC 580  
Qy 533 ACGCGTCAATTCGAACAGGTTTCACTAGCTAGCTTACGTTGCGGTTTACTCAGACTAACC 592  
Db 591 ACGCGGTGAAGTTCAGAAACGCGCGCGCTACGCTTACGCTTCAACGAGAGCGAGC 640  
Qy 593 GGTGTGTCAGGAACGTCATAGGTCGACCGGTTTCCCAAGCCATCGGTGAGCTTC 652  
Db 641 GTCTCTCGAGGAAATCTTAGCGCGCGGTTCGGAAGGCCATCGGAGCTCC 700  
Qy 653 ACGCCACACCGGTATTCGCGACTCATGTATTTACGCGCAGAGTGCAGCGGTATAG 712  
Db 701 ACGGCCACTCGCGGATCGCGCGCTCCTCTCTTACGCGCGCGTCTCTTCGCGGCTG 760  
Qy 713 TCGACCGCGCACACGAGCGGTGTAGCTAACCGCGGTTTGTCTATTTCATGCGCGGT 772  
Db 761 TTGATGGTCCAGGGATCGGCGTGGCGAAGCGCGTCTCGTCTACTTCAACACCAACC 820  
Qy 773 TATTGGCTATGTGCGAGGATTTACCTTACCAAGTTCAGATCATCTCCCAATGAGATT 832  
Db 821 TCATTGGCGATGTCGGAAGAGATTACCTTACCACGTGGAATCACCCCTTAACGCGACT 880  
Qy 833 TAAACACCGTTGTCGCTTCCGATTTGATGGACAAATAGAAATCCCAATGATGCGCCACC 892  
Db 881 TAACACCGTTGCGGTTTACGACTTCAACGGGAGCTCACTCAACAAATGATCGCCACC 940  
Qy 893 CGAAGTCGACCGGATCGGTGAATCTTCGTTTAACTTACGATCAGAGCTCGTTTCAAGC 952  
Db 941 CGAATCGGACCGCTCGACGGGACCTCCACCGCTCAGCTACGACGCTATTCAGAGC 1000  
Qy 953 CTTACCTAAATATCTTCGATTTCTTCAACGAGCGAACTAAATCACCGGAGCTCGAGATTC 1012  
Db 1001 CTTACCTCAAGTACTTCGTTTCTTCCCGCGGCTCAAGTCCCGCGAGTGGAAATCC 1060  
Qy 1013 AGCTTGATCAGCAACGATGATGACGATTTGCGGATTACAGAACTTCGTCGTGTAC 1072  
Db 1061 CCCTGAAGGACCCACCATGATGACGATTTGCGCATACGGAGATTTCTGTCGTGCTCC 1120  
Qy 1073 CTGACCAAGCTGTTTCAAGTCGCGGAGATGATCCGCGTGGGTCTCCGCTGGTTT 1132  
Db 1121 CCGACCAAGCTGTTCTTCAACTAACGAGATGATCACCGCGGTCCCCCGTGTCT 1180  
Qy 1133 ACGCAGAAACAGGTGCAAGATTCGGATTTTACAAATACGCGAAGATTCATCA 1192  
Db 1181 ACGCAAGAAACAAACCTTCAGGTTTGGGATTTGCAAGAAATGCGAGGACGCGAATG 1240  
Qy 1193 ACATTAAGTGATGATGCTCCAGATTTCTTCTGCTTCCATCTCTGGAACGCTTGGGAAG 1252  
Db 1241 CGATCGGTGATCGACGCGCGGATTTGTTCTGCTTCCACCTCTGGAACGCTGGGAG 1300  
Qy 1253 AGCAGAAACAGATGATGCTGATAGGTCCTGTATGATCTCCACAGATCATTT 1312  
Db 1301 AGCCGAAACAGAGGATTTGTGTGATTTGGTCTCTCATGATCCCTTGGGACTCCATTT 1360  
Qy 1313 TCAAGGATCTGACGAGAAATCTCAAGAGTGTCTGTCTGAAATCCGCTGAATCTCAAA 1372

Db 1361 TCAACGAATCGGAGGAGGTTTGAAGAGCGTCTGTCTCAGAGATAAGGCTGAACCTTGGA 1420  
Qy 1373 CCGGTGAATCAACTCGCGTCCGATCATCTCCAAAGAGATCAACAAGTCAACCTCGAAG 1432  
Db 1421 CCGCAAGTCCACTCGCGCCCCCATTAATCTCGACGCGC--AAAGTGAACCTGGGAG 1477  
Qy 1433 CAGGATGTTCACAGAAACATGCTCGGCGGTAAACCAAAATTCGCTTACTTGGCTTTAG 1492  
Db 1478 CCGCATGTGTGAACAGAAACAGTTCGGAAGAGACCCAGTTCGCTATCTGGCTCTGG 1537  
Qy 1493 CCGAGCGGTGCTTAAGTCTCAGGATTCGCTTAAAGTTGATCTACTACTCGAGAGTTA 1552  
Db 1538 CCGAGCGCTGCGCCAAAGTCTCGGCTTTGCGAAAGTTGATTTGCTAGTGGGAGTGA 1597  
Qy 1553 AGAAACATCTTTACGCGGATAACCGTTACGAGGAGAGCCTCTGTTTCTCCCGGAGAG 1612  
Db 1598 AGAAGTACATGTATGGAAGAGAGAGTTCGCTGGGAGGCTCTGTTTCTTCTCC--AAAG 1654  
Qy 1613 GAGGAGGAGAGAGAGGATACATCTCTGTTTCTGTTTCAAGAGAGAGACATGGAAT 1672  
Db 1655 GCCAAAGAGAGAGATGGGTATTTCTGGCATTCGTGACGAGAGAAAGATGGAAT 1714  
Qy 1673 CGGATTTACAGATAGTTAAACCGCTTAGCTTAGAGGTTGAAGCAACGTTAACTTCGT 1732  
Db 1715 CCGAGCTGAGATTTGTAATGCCAAATTTAAAGCTGAAAGTTCATCAACTCCCT 1774  
Qy 1733 CAAGGCTTCGCTACGATTTCAAGTTCAGTACATTCATCGAGCGGATTTGGCAGAGCAG 1792  
Db 1775 CTGCTTCTCTACGCTTTTATGGAATTTCTTCAATTCAGGATTTGAGAAACAG 1834

## RESULT 4

US-09-758-269-1  
; Sequence 1, Application US/09758269  
; Patent No. US20020104120A1  
; GENERAL INFORMATION:  
; APPLICANT: IUCHI, SATOSHI  
; APPLICANT: KOBAYASHI, MASATOMO  
; APPLICANT: SHINOZAKI, KAZUO  
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN  
; FILE OF INVENTION: CLEAVAGE ENZYME GENE  
; FILE REFERENCE: 3914-3  
; CURRENT APPLICATION NUMBER: US/09/758,269  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: JP 2001-003476  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: JP 2000-010056  
; PRIOR FILING DATE: 2000-01-13  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1752  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1749)  
US-09-758-269-1

Query Match 41.0%; Score 738.2; DB 10; Length 1752;

Best Local Similarity 68.2%; Pred. No. 9.8e-244; Indels 18; Gaps 2;  
Matches 1063; Conservative 0; Mismatches 478;

Qy 244 AACACTAAACAGATGAATTTGTTCCAGAGAGCGCGGCGGAGCGTTGAGACCGCGGAG 303  
Db 193 AATCTCTCCGGCTAAACATCTTCCAGAAAGCGCGGAGATTGCGATCGACGCGCTGAG 252  
Qy 304 GGTTCCTTGTGACGACGAGAGCTACACCGCTTCTTAAACGCTGATCTTAGTGT 363  
Db 253 CGTGCTAATATCTCACACGAGCAAGATTCTCACATTTCCAAACCGCTGATCCAGTGT 312  
Qy 364 CAGATCGCCGAAATTTGCTCGGTGAATGAACAGCCCGCTCCGCGTAAATCTTCCGCTG 423

Db 313 CAGATTCCGGGAATTATTCCTCCGGTACCGGAATCTTCCGTCCGGCGAAACCTCACCGTC 372  
QY 424 GTCCGAAAATTCCCGATTCATCAAGAGGATGATGTGGCAACGAGCTAACCCACTT 483  
Db 373 GAAGGAACAATCCCTGATGCAATGACGGTGTATATATCCGTAAACGGCGGAATCCGATG 432  
QY 484 CACGAGCCGGTGACAGGTGACCACTTCTTCACGGAGACGGTATGTTTCACGCCGTCAAA 543  
Db 433 TTTGAGCCAAACAGCTGGGCACCAATTTATTCAGCGAGACGGAAATGGTTTCACGCGATTAA 492  
QY 544 TTCGAAACGGTTTACGCTAGCTTGCCTGCTTACTCAGACTAACCGGTTTGTTCAG 603  
Db 493 ATAACCAACGGTTTACGCTAGCTAGCTAGCTGCTGCTTACAAAACCGAGATTGGTTTCAG 552  
QY 604 GAACTCAATTTGGTTCGACCGGTTTCCCAAGCAATCGGTGAGCTTCACGGCCACACC 663  
Db 553 GAAACACGATTTGGTTCGACCGGTTTCCGAAAGCAATCGCGAGCTTCACGGTCACTCG 612  
QY 664 GGTATGTCGCACTCATCTATCTACGCCAGAGCTGCGAGCGGTATAGTCGACCCGGCA 723  
Db 613 GGAATCGCACGTTGATCTGTTTACGCACGTGGGCTTTTGTGCTCTGATCAACAACCAA 672  
QY 724 CACGGAACGGGTAGCTTAACCGGTTTGGTCTATTTCAATGCCGTTATTCGCTATG 783  
Db 673 AACGGCTCGAGTAGCAACGCCGGTTTGGTTTACTTTAATAACCGGCTTTTAGCTATG 732  
QY 784 TCGGAGGATGATTTACCTTTACCAAGTTCAGATCACTCCCAATGGAGATTTAAAAACCGTT 843  
Db 733 TCAGAAGACGATTTACCTTACCAATTAATAAATTACTCAACCGCGATCTCCAAACCGTT 792  
QY 844 GGTGGTTTCGATTTTGGTGGCAATTAAGATCCCAATGATGCTCCCGGAAAGTTCGAC 903  
Db 793 GGAAGTTTACGATTTTCGAGCGTCAATTAATCCGCAATGATGCTACCCGAAACTGGAC 852  
QY 904 CCGGAATCCGGTGAATCTTCCGTTTAACTACGACGCTGCTTCAAGACCTTACCTAAA 963  
Db 853 CCGGTTACGAGGAGCTTCACGCTTAAGCTACGAGCTGTTAAGAAACCTTACCTGAA 912  
QY 964 TACTTCCGATTTTCACCGGAGGAACTAAATCAACGGAGCTGAGATTCAGCTTATGATCAG 1023  
Db 913 TACTTTCAGATTTTCGCCAGAGCGGCTTAAATCGCGGAATGGAGATCCCGCTCGAACT 972  
QY 1024 CCAACGATGATCACGATTTCCGATTTACAGAGAACTTCGCTGCTGATCTGACGACAA 1083  
Db 973 CGACGATGATTCAGATTTCCGATTAACGGAGATTTTGGTGATTCCTGATCAACAA 1032  
QY 1084 GTCTGTTTCAAGCTCCGAGAGATGATCCGGGTGGGTCTCCGGTGGTTTACGACAGAAC 1143  
Db 1033 GTCTGTTTCAAGCTCCGAGAGATGATTTCCGGTAAATCTCCGGTGTGTTTTCGACGGGAA 1092  
QY 1144 RAGGTCCGAAGATTCGGGATTTTACGAAATACGCGGAAGATTCATCGAACTTAAGTGG 1203  
Db 1093 AAGGTTCCCGATTTGGGATTAATGCCAAGACGCGACAGAGCTTCTCAGATTAATCTGG 1152  
QY 1204 ATTGATGTTCCAGATTTGCTGCTTCATCTCTGGAACCGTTGGGAAGAGCCGAGAAACA 1263  
Db 1153 GTGAATCTCTCCGAGACGTTCTGTTTTCATCTCTGGAATGCAATGGAATCCCGGAGACG 1212  
QY 1264 GATGAAGTCTGCTGATAGGTCCTGTATGATCTCCACGAGCTCAATTTTCAAGAGTCT 1323  
Db 1213 GAGGAGATTTGGTGAATCGATGCTGTATGTCGCGGGGATTAATCTTCAACGAGAGA 1272  
QY 1324 GACGAGAATCTCAAGAGTGTCTGCTGAAATCCGCTTGAATCTCAAAACCGGTGAATCA 1383  
Db 1273 GACGAGAGCTTGAAGCGTTTGTTCGAGATCAGATAAACCCTCAGAACACGCTAAACCC 1332  
QY 1384 ACTCGCCGTCGATCATCTCCAAAGAGATCAACAGTCAACCTCGAAGCAGGATGGTC 1443  
Db 1333 ACGGCTGCTGCTGTTGTTTGAACGAGGAT-----GTAATTTAGAGATTGGTATGGTT 1386  
QY 1444 AACAGAAAATGCTCGCCCGCTAAAACCAAAATTCGCTTACTTGGCTTTTAGCCGAGCCGTGG 1503  
Db 1387 AACCGGAACCGGTTAGGAGAGAAACCCCGGTTTCGCGTTTTCGCTTATGCTTATCTCTGG 1446

QY 1504 CCTAAAGTCTCAGGATTCGTTAAAGTTGATCTCACTACTGAGAAAGTTAAGAAAATCTT 1563  
Db 1447 CCAAAAGTTTCGGTTTCGTTAAAGTTCGATCTTTTCACCGGTGAGATGAAAATATATT 1506  
QY 1564 TACGGCGATAACCGTTTACGGAGGAGGCTCTGTTTCTCCCGG-----AGAA 1611  
Db 1507 TACGGCGGTGAGAAATATGGCGCAACCGTTTCTTCCCGGCAACTCCGGTTAACCGC 1566  
QY 1612 GGAGGAGAGGAGACGAAGGATACATCTCTGTTTTCGTTTTCAGACGAGAACATCGGAAA 1671  
Db 1567 GAAGAAATGAAGATGACGGTTTATATATTTGTACGTTTCATGACGAGAAACAAGACA 1626  
QY 1672 TCGGAGTTACAGATAGTTAAACCGCTTACGTTAGAGTTGAAGCAACGGTTAACTTCG 1731  
Db 1627 TCAGAGCTTCAGATTAATAACGCTGTAAATTTAAAGCTTGAAGCTACGATTAAACTACCG 1686  
QY 1732 TCAAGGTTCCGTACGGAATTCACGTTACATTCGAGCGCGATGATTTGCGGGAAGCA 1790  
Db 1697 TCTAGAGTACCGTATGGGTTTCTGCGACATTTGGGATTCGATGAATCGTCTGATCA 1745

## RESULT 5

US-09-758-269-13  
; Sequence 13, Application US/09758269  
; Patent No. US20020104120A1  
; GENERAL INFORMATION:  
; APPLICANT: IUCHI, SATOSHI  
; APPLICANT: KOBAYASHI, MASATOMO  
; APPLICANT: SHINOZAKI, KAZUO  
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN  
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE  
; FILE REFERENCE: 3914-3  
; CURRENT APPLICATION NUMBER: US/09/758,269  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: JP 2001-003476  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: JP 2000-010056  
; PRIOR FILING DATE: 2000-01-13  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 1815  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1812)  
US-09-758-269-13

Query Match 33.7%; Score 606.6; DB 10; Length 1815;  
Best Local Similarity 63.5%; Pred. No. 3.5e-198;  
Matches 1003; Conservative 0; Mismatches 549; Indels 27; Gaps 4;

QY 246 CACTAACAGATGAATTTGTTCCAGAGAGCGCGCGGCGGCTTGGACGCGCGGAGG 305  
Db 240 CAAGAAGCAGCTCAACTTGTTCAGCGCGCGCGCGCGCGCTCGACGCGTTCCAGGA 299  
QY 306 TTTCTCTGT-----CAGCCACGAGAAGCTACACCCGCTTCTTAAACCGGCTGATCCTAG 359  
Db 300 AGGTTTCGTGCCAACGTCCTCGAGCGGCCCGACGCGCTGCCAGCACGCGCGACCCGGC 359  
QY 360 TGTTTCAGATCGCGGAAATTTTCTCGGTGATGAAACACCGCGCTCCGGGTAACTTTC 419  
Db 360 CGTTCAGATCGCGGAACTTTCGCGCCCTTCGCGGAGAGCGCGCGCTGACGAGTCTCC 419  
QY 420 GGTGTTCCGAAAATCTTCCCGATTCCTCAAGAGAGTGTATGTCCGCAACGAGCTAACCC 479  
Db 420 CGTCTCGCGCGCATCCGCCCTTTCATCGACGGGTCTACGCGCGCAACGCGCAACCC 479  
QY 480 ACTTCAGAGCGGTGACAGGTCAACCTTCTTCGACGAGAGCGGTATGGTTCAACCCGT 539  
Db 480 CTGCTTCGACCCCGTCCGCGGCGACCACTCTTCGACGCGCGACGCAATGTTGACGCGCT 539

QY 540 ---CAAATTCGAACACGGTTGAGCTAGCTACGCTTGCCTGTTTACTCAGACTACCGGTT 596  
Db 540 GGGGATACGAACGGCGCGGAGTCTAGCCTTGCCTTACGGGACCGCGCGCT 599  
QY 597 TGTTCAGAAACGTCAATTTGGTTCGACGGTTTCCCAAAGCCATTCGGTTCAGCTACGG 656  
Db 600 GCGCAGGAGCGCGGATCGCGCGCGCTTTCGCCAAGGCCATTTGGCGAGCTGACCG 659  
QY 657 CCACACCGGATTGCGCGACTCATCTATTCTACGCCAGAGTCGAGCGGTATAGTCGA 716  
Db 660 GCACTCCGGATTCGCGCGCTTCCCTGTTCTACGCGCGCGGTGCGGCTCTGTCGA 719  
QY 717 CCGGACACACGGAACCGGTGAGCTACGCGGTTTGGTCTATTTCATTCGCGGTTATT 776  
Db 720 CCGCTCGCGCGCACCGCGGTGGCCAAACGCGCGCTCTACTTCAACGCGCGCGCTGCT 779  
QY 777 GCGTATGCGGAGGATGATTTACCTTACCAAGTTTCAGATCACTCCCAATGGAGATTAAA 836  
Db 780 CGCCATGTCGAGGACGACCTCCCTTACCAGTCCGCGTGGGAGCGAGCGGACCTCGA 839  
QY 837 AACCGTGGTTCGGTTTCGATTTTGATGGCAATTAGAATCCCAATGATTTGCCACCCGAA 896  
Db 840 GACCGTTCGCGCTACGACTTCGACGGGAGCTCGGCTCGGCTATGATCGCGCACCCAA 899  
QY 897 AGTCGACCGGAAATCCGGTGAATCTTCCGCTTAACTACGACGCTGTTTCAAGCCCTTA 956  
Db 900 GTGGACCGCGCACCGGGAGTCCACGCTTCAGCTACGACGCTATCAAGAGGCGCGTA 959  
QY 957 CTTAAATATCTCCGATTTCTACCGGACGGAATTAATCACCGGACGTCGAGATTCAGCT 1016  
Db 960 CTTCAAGTACTTCTACTTCAGGCCGACGCGACCAAGTCCGACGAGCTGGAGATCCGCT 1019  
QY 1017 TGATAGCCAAACGATGATGACGATTTCCGATTCAGAGAACTTCGCTGCTACCTGA 1076  
Db 1020 GGAGCAGCCACGATGATCCAGACTTCGCCATCCACCGAGAACTTCGTTGTTTGCCTGA 1079  
QY 1077 CCAGCAAGTCGTTTCAAGCTGCGGAGATGATCCGCGTGGTCTTCGCTGTTTACGA 1136  
Db 1080 CCACAGTGTGTTTCAAGCTTCAGAGATGCTGCGCGGGTTCGCGCTGCTGCTGA 1139  
QY 1137 CAAGAACAGGTCGCAAGATTCGGGATTTTAGAACAATACGCGGAGATTCATCAACAT 1196  
Db 1140 CAAGGAGAACGCTCGCGTTCGGCTGCTCCCAAGCACGCGCGGACGCTCGGAGAT 1199  
QY 1197 TAAGTGATGATGCTCCAGATTTCTGCTTCCATCTCTGGAACGCTTGGGAAGCGC 1256  
Db 1200 GCGTGGTGGAGCTGCGCGACTGCTTCTGCTTCCAGCTTGAACGCTGGGAGACA 1259  
QY 1257 AGAAACAGATGAAGTCGTGATGAGTTCCTGATGATCTCCACGAGATCAATTTCAA 1316  
Db 1260 GCGACCGGCGAGTGGTGTGATCGGCTCTGATGACCCCGCGACTCCATCTTTCAA 1319  
QY 1317 CGAGTCTGACGAGATCTCAGAGTTCCTGCTGAAATCCGCTGAAATCTCAAAACCGG 1376  
Db 1320 CGAGTCCGACGAGCGCTTGAGAGCGTGTCTGACGAGATCCGCTGAGCGCGCACGG 1379  
QY 1377 TGAATCAACTCGCGCTCCGATCATCTCCAAACGAGATCAACAAGTCAACCTCGACAGG 1436  
Db 1380 CCGCTCCACGCGCGCGCTCTGCGC---GCGCTCGCAGCAGGAGAACTCGAGGTGGG 1436  
QY 1437 GATGTCACAGAAACATGCTCGCGCTTAAACCAATTCGCTTACTTGGCTTTACCGA 1496  
Db 1437 CATGTCGACCGCAACCTGTGGCGCGGAGAGCGCGTACGCTACCTCGCGTGGCGA 1496  
QY 1497 GCGTGGCTTAAAGTCTCAGGATTCGCTAAAGTTGATCTCACTACTGGAGATTTAGAA 1556  
Db 1497 GCGTGGCGCAAGGAGTGGGCTTCGCCAAGGAGGACCTGTCCACGGCGAGCTCACCA 1556  
QY 1557 ACATCTTTACGCGATACGCTTACGAGGAGGAGCTCTGTTCTCCCGGAGAGGAG- 1615  
Db 1557 GTTCAGTACGCGGAGGCGGTTTCGCGCGGAGCCCTGCTTCTGTTCCCATGACCGCGC 1616

QY 1616 -----GAGAGGAAGACGAAGGATATATCTCTCTCTCTCTCTCAGACGAGAA 1661  
Db 1617 CGCGGCCACCGCGCGCGAGGACGAGGTTACGTCTCCTCTCTCTCTCAGACGAGCG 1676  
QY 1662 GACATCGAATCGGAGTTACAGATAGTTAAACCCGCTTAGCTTAGAGGTTGAAGCAACCGT 1721  
Db 1677 CGCGGACAGCTCGGAGCTACTTGTGTCNAATGCGCGGACATCCGCTGGAGCGCACGCT 1736  
QY 1722 TAAATTCCTCAAGGTTCCGTACGATTTTCAAGTATCAATTCATCGGAGCGGATGATT 1781  
Db 1737 TCAGTTCGCTCCGCGTCCGCTTCGCTTCCACGCGACCTTCATCAGCGGCGGAGGCT 1796  
QY 1782 GCGAAGCAGGTCGTGTA 1800  
Db 1797 CGAGGCCAGGCGGCTGA 1815

## RESULT 6

US-09-758-269-9  
; Sequence 9, Application US/09758269  
; Patent No. US20020104120A1  
; GENERAL INFORMATION:  
; APPLICANT: IUCHI, SATOSHI  
; APPLICANT: KOBAYASHI, MASATOMO  
; APPLICANT: SHINOZAKI, KAZUO  
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN  
; FILE REFERENCE: 3914-3  
; CURRENT APPLICATION NUMBER: US/09/758,269  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: JP 2001-003476  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: JP 2000-010056  
; PRIOR FILING DATE: 2000-01-13  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 1734  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1731)  
US-09-758-269-9

Query Match 29.3%; Score 528.2; DB 10; Length 1734;  
Best Local Similarity 61.0%; Pred. No. 4,7e-171;  
Matches 920; Conservative 0; Mismatches 563; Indels 24; Gaps 3;

QY 306 TTTCCTTTGTCAGCCACGAGAGCTACACCGCTTCTTAAACCGCTGATCTAGTGTCA 365  
Db 240 TATCGTTATTCCTATGGAGCAGATCCCGCTTCTTAAACCGACCGCGCGGTCA 299  
QY 366 GATCGCGGAAATTTCTCGGTGAATGAACAGCCGCTCCGGGTAAATCTTCGGTGGT 425  
Db 300 ATTATCAGGTAACTTCGCTCGGTAAATGAATTCGCGTTTCAAGATGGT 359  
QY 426 CGGAAACTTCCGATTCATCAAGGAGTGTATGTCGCAACGAGCTAACCCACTCA 485  
Db 360 TGGTCAGATTCCTTCTGTCTAAAGGAGTTTATCATCCGTAAACCGTAAACCTTGT 419  
QY 486 CGAGCGGTGACAGTCAACCACTTCTTCGACGGAGACGGTATGTTTCAACCGTCAATT 545  
Db 420 TCCGCGTTAGCGGACATCATTTTATTCGCGTACGGAATGATTCACGCGTTAGTAT 479  
QY 546 CGAACACCGT---TCAGTAGTACGTTGCGGTTTACTCAGACTAACCGTTGTTC 602  
Db 480 CGGTTTGTATAACACGAGTTAGTTACGTCGCGGTACACTTAAACAAACCGCTTGTTC 539  
QY 603 GGAAGCTCAATTTGGTTCGACCGGTTTCCCAAGCCATCGGTGAGCTTCAACGCCAC 662  
Db 540 AGAACCCGCGCTTGAGCGATCGGTTTTCCTTAAACCAATCGCGAGCTTCAACGCCATC 599



QY 663 CGGTATTCGCCGACTCATGCTATTCTAGCCGAGAGCTGACGCGGTATAGTCGACCCGGC 722  
Db |||||  
QY 600 CGGTCTAGCTCGACTCGCTCTCTTACCGCTGAGCTGGGATCGGTCTAGTGGACGGGAC 659  
Db |||||  
QY 723 ACACGGAAACCGGTGATAGCTAAACCGGTTTGGTCTATTATTCATGCGCGGTTATTGGGTAT 782  
Db |||||  
QY 660 AGTGGCATGGCGTAGCTAACCGCGGTGGTCTTCTTTAACGCGAGTTATTAGCCAT 719  
Db |||||  
QY 783 GTCGAGAGATGATTTACCTTACCAATTCAGATCACTCCCAATGGAGATTTAAAAACCGT 842  
Db |||||  
QY 720 GTCAGAAGATGATCTCTCTTACCAAGTGAAGATCGACGCTCAAGGAGATCTTGAGACGAT 779  
Db |||||  
QY 843 TGGTTCGGTTCGATTTTGTATGGACAATTAGAAATCCCAATGATTTGCCACCCGAAAGTCGA 902  
Db |||||  
QY 780 CGAGCGGTTCCGATTCGATGACCAATTCAGTCTTTCAGTGATAGCGCATCTTAGGTGA 839  
Db |||||  
QY 903 CCGGAATCCGGTGAACCTTCGCTTTTAAGCTTACGACGTCGTTTCAAGCGCTTACCTTAA 962  
Db |||||  
QY 840 CGCGACCAAGGAGATCTCCATACACTGAGCTACCAACGTTTGAAGAAACCTCATCTCAG 899  
Db |||||  
QY 963 ATACTTCGATTTCTACCGGAGGAACTAAATCACCGGAGTTCGAGATTCAGCTTGATCA 1022  
Db |||||  
QY 900 GTATCTTAAATTCACACGTGGGAAAGACACGTGACGTGGAGATCACGCTCCCTGA 959  
Db |||||  
QY 1023 GCCAAGATGATGACGATTTCCGATTTACAGAGAACTTCGTCGTCGACCTGACCAACA 1082  
Db |||||  
QY 960 ACCAACGATGATTCATGATTTCCGATTAACCGAGATTTTGTGCTTATACCGGATCAGCA 1019  
Db |||||  
QY 1083 AGTCGTTTCAAGCTCCCGAGATGATCCGCGTGGGTCGCGTGGTTCAGGACAGAA 1142  
Db |||||  
QY 1020 AATGGTATTCAAATTCATCGAATGATTCGGGCGGGTCACCCGTTATCTAGTTAAAGA 1079  
Db |||||  
QY 1143 CAAGGTCGCAAGATTCGGGATTTTAGCAAAATACCGGAGATTCATCGAATTAAGTG 1202  
Db |||||  
QY 1080 AAAATGGCGAGATTTGGAGTTTGTCAAGCAGGATCTGACCGGTTCGATATAAATG 1139  
Db |||||  
QY 1203 GATTGATGCTCAGATGCTTCGCTTCCATCTCTGGAACGCTTGGAGAGCCGAGAAC 1262  
Db |||||  
QY 1140 GGTGATGATACCGGATTTGCTGTTTCCATCTATGGAATGCGTGGGAGAGAGAACCGA 1199  
Db |||||  
QY 1263 AGATGAAG-----TCGTGTGATAGGTCCTGTATGATCTCCACAGACTCAATTT 1313  
Db |||||  
QY 1200 AGAGGAGAGCCAGTTATGCTGTATCGGTCAITGATGAGCCCAACCGACAGCTTT 1259  
Db |||||  
QY 1314 CAACGAGTCTGACGAGAAATCTCAAGAGTGTCTGTCTGAAATCCGCTTGAATCTCAAAC 1373  
Db |||||  
QY 1260 TAGTGAATCAGGAGAACCAACCGGTTGAAATTAAGTGAGATCCGCTTAAACATCGTAC 1319  
Db |||||  
QY 1374 CGGTGAATCAACTCGCGTCCGATCATCTCCAAGAGATCAACAGTCAACCTCGAAGC 1433  
Db |||||  
QY 1320 AAAAGATCGAACCGTAAAGTTATCGTAACC-----GGAGTGAATTTAGAGC 1367  
Db |||||  
QY 1434 AGGATGGTCAACAGAAACATGCTCGGCGGTAAACCAATTCGCTTACTTGGCTTTAGC 1493  
Db |||||  
QY 1368 GGGTCAATAAACCGTAGTTACGTGGGCGGAAAGCCAGTTGCTTTTACATAGCAATAGC 1427  
Db |||||  
QY 1494 CGAGCGTGCCTTAAAGTCTCAGATTCGCTAAAGTTGATCTCACTACTGGAGAGTTAA 1553  
Db |||||  
QY 1428 CGATCTCTGGCCCAATGCGAGTGGCATTCGGAAGGTAGATATACAAACCGCACCGTTTC 1487  
Db |||||  
QY 1554 GAAACATCTTTACCGGATTAACCGTTACGGAGAGAGCCCTCTGTTTCTCCCGGAGAGG 1613  
Db |||||  
QY 1488 AGAGTTTAAATTCAGGACCGAGCCGTTTCGGTGAGAAACCGTGTCTTGTACCGGAGGAGA 1547  
Db |||||  
QY 1614 AGGAGAGAGACGAGGATACATCTCTGTTTCGTTACGAGGAGAACATGGAATC 1673  
Db |||||  
QY 1548 AGGAGAGAGAGACAAAGGTTATGATGCGGTTTGTGAGAGAGACGAGAGAGACGAGTC 1607  
Db |||||  
QY 1674 GGAATTACAGATAGTTAAACCGGTTAGCTTAGAGGTTGAGAGCAACGTTTAACTTCGTC 1733  
Db |||||  
QY 1608 GGAGTTTGGTGGTGCAGCCGAGGATATGAAGCAAGTCGCGCGGCTGCGCTTGCCGA 1667  
Db |||||  
QY 1734 AAGGTTCCGTAACGATTTACCGGTATCATTCATCGGAGCCGATGATTTGGCGNAGCAGT 1793  
Db |||||

RESULT 8

Db 1668 GAGGTACTTATGTTTCCATGGAACTGTTCTGAGCGAGATCAGTTGAAGRAAAGT 1727  
QY 1794 CGTGCA 1800  
Db 1728 TTTCTGA 1734

RESULT 7

US-09-770-696-172/c  
; Sequence 172, Application US/09770696  
; Patent No. US20010044940A1  
; GENERAL INFORMATION:  
; APPLICANT: Gorlach, Jorn  
; APPLICANT: An, Yong-Qiang  
; APPLICANT: Hamilton, Carol M.  
; APPLICANT: Price, Jennifer L.  
; APPLICANT: Raines, Tracy M.  
; APPLICANT: Yu, Yang  
; APPLICANT: Rameaka, Joshua G.  
; APPLICANT: Page, Amy  
; APPLICANT: Matthew, Abraham V.  
; APPLICANT: Ledford, Brooke L.  
; APPLICANT: Woessner, Jeffrey P.  
; APPLICANT: Haas, William David  
; APPLICANT: Garcia, Carlos A.  
; APPLICANT: Krickler, Maja  
; APPLICANT: Slader, Ted  
; APPLICANT: Davis, Keith R.  
; APPLICANT: Allen, Keith  
; APPLICANT: Hoffman, Neil  
; APPLICANT: Hurban, Patrick  
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
; FILE REFERENCE: 2031US (PAPA-020PRV)  
; CURRENT APPLICATION NUMBER: US/09/770,696  
; CURRENT FILING DATE: 2001-01-26  
; PRIOR FILING DATE: 2000-01-27  
; NUMBER OF SEQ ID NOS: 911  
; SOFTWARE: Fast-SEQ for Windows Version 4.0  
; SEQ ID NO 172  
; LENGTH: 200  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(200)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-770-696-172

Query Match 11.0%; Score 198; DB 9; Length 200;  
Best Local Similarity 99.0%; Pred. No. 2.1e-57;  
Matches 198; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 792 TGATTTACCTTACCAAGTTCAGATCACTCCCAATGGAGATTTAAAAACCGTTGTCGGTT 851  
Db 200 TGATTTACCTTACCAAGTTCAGATCACTCCCAATGGAGATTTAAAAACCGTTGTCGGTT 141  
QY 852 CGATTTTGAAGCAATTAATCCCAATGATGCCCCCGGAAAGTCGACCCGGAATC 911  
Db 140 CGATTTTGAAGCAATTAATCCCAATGATGCCCCCGGAAAGTCGACCCGGAATC 81  
QY 912 CGGTGAATCTTCGCTTTAAGCTACGAGCTCGTTTCAAGCCCTTACCTAAAAATCTTCG 971  
Db 80 CGGTGAATCTTCGCTTTAAGCTACGAGCTCGTTTCAAGCCCTTACCTAAAAATCTTCG 21  
QY 972 ATTCTACCGGACGGAACCTA 991  
Db 20 ATTCTACCGGACGGAACCTA 1

RESULT 8

```

US-09-878-574-2543
; Sequence 2543, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J. D.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 2543
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-023-Q1-B1-G10
US-09-878-574-2543

Query Match 10.6%; Score 190.4; DB 10; Length 393;
Best Local Similarity 68.5%; Pred. No. 1.4e-54;
Matches 263; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 981 GGACGGAACTAATCACCGGACGTCGAGATTCAGCTTGATCAGCCACGATGATGCACGA 1040
DB 7 GGACGAGAGAAAGTCGCGGACATAGAAATTCCTTCGCGCGGACGATGACGACGA 66
QY 1041 TTTCGGGATACAGAGAACTTCGTCGTCGTACCTGCACGACGAGTCGTTTCAAGTGGC 1100
DB 67 CTTCCGGATCACGGAGATTCCTGGTGATCCCGGACCAGCAGGTGTTTCAAGTCGG 126
QY 1101 GGNAGTATCCGGTGGTCTCCGGTGGTTTACGACGAGAACAGTCCGAGATTCGG 1160
DB 127 CGAAATGATCAAGGAGGATCCCGGTGATCTACGCGCGGAGAGAAATCGCGTTCGG 186
QY 1161 GATTTTAGACAAATACGCCGAGATTCATCGAACATTAAAGTCGATTGATCTCAGATTG 1220
DB 187 AATACTGCCAAAGTACGCTTCGACGCGCTCAGCATCTGTGGTGGACTCGGCAGACAC 246
QY 1221 CTTTCGCTTCATCTCTGGAACGCTTGGGAGAGCCAGAACACATCAAGTCGTCGTGAT 1280
DB 247 GTTCTCTTCCACTTCTGGAACGCTGGGAGGACCGCAACGAAATGTTGGTGAAT 306
QY 1281 AGGTCCTGTATGACTCCACCAAGACTCAATTTTCAACGAGTCTGACGAGAATCTCAAGAG 1340
DB 307 AGGTTCTGTCATCACACCACCAAGATTCATTAATCACAGAGAGAGACTTGAAGAG 366
QY 1341 TGTCTCTCTGTAATCCGCTGAA 1364
DB 367 CGTTTAAAGAAAGTAAGCTGAA 390

RESULT 9
US-09-758-269-3
; Sequence 3, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33

```

QY 1311 TTTCACGAGTCTGACGAGAAATCTCAAGAGTGTCTCTGTAATCCGCCTGATCTCAA 1370  
DB 1290 TTGAGAGGATGGATCTGGTTC---ATGCTTTGGTGGAGAGTGAAGATCGATCTCGT 1346  
QY 1371 AACCGGTGAATCAATCCGCGCTCCGATCATCTCCAACGAAGATCAACAAGTCAACCTCGA 1430  
DB 1347 CACCGGATGTGACAGCTCATCCGATCTCAGCGAGG-----AATCTCGA 1391  
QY 1431 AGCAGGATGTCAACAGAAATCTCTCGCGGTAAACCAATTCGCTTACTTGGCTTT 1490  
DB 1392 TTTCCTGTGATTAATCCGGGTTTCTCGGAGATGTAGCAGGTACGTTTACGCGGGAT 1451  
QY 1491 AGCCGAGCGGTGGCTAAAGTCTCAGGATTCGCTAAAGTTCATCTCACTACTGAGAGAT 1550  
DB 1452 TGAAGATCCGATGCCGAGATCTCCGCTGTGTGAGCTTGTATGTCTAAAGAGATCG 1511  
QY 1551 TAAGAAACATC-----TTTACGCGATTAACCGTTACGAGAGAGCCTCT 1595  
DB 1512 GGATGATTTAGCGTGGCGGTGAGATGTACGGTTTACGCGGAGAACCGTT 1571  
QY 1596 GTTCTCCCGAGAGAGGAGG-----GAGGAAGACGAAGGATACATCTCTG 1643  
DB 1572 TTTCGTAGTAGGATCTCTGTAAATCCGAGCGGAGGAGATGATGTTATGTGTGAC 1631  
QY 1644 TTTCCTTCAACGACGAGAGACATGAAATCGGAGTTACAGATGTAACGCGTTAGCTT 1703  
DB 1632 GTATGTTACGATGAAGTACTGAGAGATCGAAGTTTCTGTGTATGAGCGCTAAATCGCC 1691  
QY 1704 AGAGTTGAA-----GCAACGGTTAAACTTCCGTCAGAGTTCCGTCAGGATTTACGG 1757  
DB 1692 GGAGCTTGAATCTGTCGCGCGCTGAGGTTGCGCGAAGGTTCCGTACGATTCCTATGG 1751  
QY 1758 TACATTCATCGAGCGGATGATTTGCGGAAGC 1789  
DB 1752 GTTATTTGTCAGGAAGTGAACCTTAATAAGC 1783

## RESULT 10

US-09-938-842A-1444  
; Sequence 1444, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; FILE OF INVENTION: SAME, AND METHODS OF USE  
; FILE REFERENCE: SCRI1300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; CURRENT FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 1444  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-1444

Query Match 9.4%; Score 169.6; DB 10; Length 1788;

Best Local Similarity 50.2%; Pred.No. 6.1e-47; Matches 709; Conservative 0; Mismatches 634; Indels 69; Gaps 9;

QY 426 CGGAAACTTCCGATTCATCAAGAGGATGTATGTGCGCAACGAGGTAAACCACTTCA 485  
DB 393 CGGCACTTTCCTGCTCACTTAACGGCGTTACATCCGTAACGGTCCAAATCCACACT 452

QY 486 CGAGCGGTGACAGGTACACCACTTCTTGACGAGACGGTATGGTTTACGCGGTCAAAT 545  
DB 453 TCTCCCTCTGTTGCTTACCATCTCTTGACGCGGACGGTATGCTTTCAGCCATAAAT 512  
QY 546 CGAACACGGTTACGTTAGCTAGCTTACGTTGCGGTTTACTAGATAACCGGTTGTTTCA 605  
DB 513 CCACAAACGGTAAAGCCACTCTCTGTAGCAGATACGTCAAGACTTATAAATACAACGT 572  
QY 606 ACCTCAATTTGGTTCGACCGGTTTCCCAAAGCCATCGGTGAGCTTCAACGCGCACACCG- 664  
DB 573 GAAACAAACCGGAGCTCCGGTTATGCTTAACGTTTCCGGATTCACCGGTGTAACGCG 632  
QY 665 --GTATTGCCGACTCATCTATTCTACGCCAGAGTGCAGCGGTATAGTGCAGCCGCG 722  
DB 633 GTCAGTAGCTCGTGGAGCTTTTAAACGCACTAGGTTTAAACGCGACAGTATAATCCGT 692  
QY 723 ACACGAAACCGGTAGCTAAACGCGGTTTGTCTATTTCAATGCGCGGTTATTTGCTAT 782  
DB 693 TAAACGCAATTTGGTTAGCTAATAAAGTCTAGCTTTCTTCACTAACCGTCTCTTTGCT 752  
QY 783 GTCGAGGATGATTTACCTTTACCAAGTTCCAGATCACTCCCAATGGAGATTTAAAACCGT 842  
DB 753 AGGTGAATCTGATTTTACCCTAGCCCTCGATTAACCGAATCAGGAGATATTGAACGAT 812  
QY 843 TGTCTCGTTGATTTTGTATGGACATTAAGATTCACCAATGATTCGCCACCGGAAAGT 902  
DB 813 CGACCGGTAGATTTTCGACGGGAAATTTAGCGATGATGACAGCTCATCTTAAACCGA 872  
QY 903 CCGGAAATCCGTTGAACTCTTTCGCTTTTAAAGCTACGACGTCGTTTCAAAGCCTTACCT 962  
DB 873 TCCAATAACCGGAGAAACTTTGCTTTCCGTTACGCTCCGTTTCCA---CCGTTTTTAA 929  
QY 963 ATACTTCCGATTTCTACCGGACGGAATTAATCAACCGGACGTCGAGA---TTCAGCTTGA 1019  
DB 930 ATATTTCCGTTTGTATTCGCGCGGAAACAAAGAGACGTTCCGATATCTCGATGAC 989  
QY 1020 TCAGCCAAACGATGATGACGATTTCCGATTTACAGAGAACTTCGCTCGTACTGACCA 1079  
DB 990 GTCTCGTCTGTTCTCCATGACTTCCGATCAGAAACGTCACGCGATTTTCGACAGAT 1049  
QY 1080 GCAAGTCG-----TTTTCAAGCTCGCGGAGATGATCCGCGTGGGTCTCCGTTGGT 1130  
DB 1050 TCAGCTTGGCATGAGGATGAACATGTTGGATTGTTCTCGAAGTGGTCTCCGTTGG 1109  
QY 1131 TTACGACAAACAAAGTCCGAGATTCGGGATTTTACAAATACGCGAAGATTCATC 1190  
DB 1110 TACTGATACGGAATAACTCCAAAGCTTGGAGTATCTTAAGTACCGCGAGATGATC 1169  
QY 1191 GAAATTAAGTGAATGATGCTCCAGATTCGTTCTGCTTCCATCTCTGAAACGCTTGGGA 1250  
DB 1170 GGAGATGAAATGGTTCGAAGTTCCTGATTCATATCAATTCAGCTTATTAATGCTTGGGA 1229  
QY 1251 AGACCCAGAAACAGATGAGTCTGATAGGTCCTGATGAGTCCCTGATGACTCCACGACTCAAT 1310  
DB 1230 TGAAGATGATGGAACAGCGCTCTTTGATTCACCGAATATTTATGCTGATGAAATAC 1289  
QY 1311 TTTCACGAGTCTGACGAGAACTCAAGAGTGTCTGTCTGAAATCCGCTGAAATCTCAA 1370  
DB 1290 TTTAGAGAGGATGGATCTGTTTC---ATGCTTTGGTGGAGAGGTGAAGATCGATCTCGT 1346  
QY 1371 AACCGGTGAATCAACTCCGCTCGGATCATCTCCAACGAGATCAACAGTCAACCTCGA 1430  
DB 1347 CACCGGATTTGAGAGCTCATCCGATCTCAGCGAGG-----AATCTCGA 1391  
QY 1431 AGCAGGATGTCACAGAAACATGCTCGGCCCTTAAACCAATTCGCTTACTTGGCTTT 1490  
DB 1392 TTTGCTGTGATTAATCCGCGGTTTCTCGGAGATGTAGCAGGTACGTTTACGCGGAT 1451  
QY 1491 AGCGAGCGGTGGCTTAAAGTCTCAGGATTCGTTAAAGTTGATCTCACTACTCGGAGAT 1550  
DB 1452 TGGAGATCCGATCCGAGATCTCCGTTGTGGTGAAGCTTGTGTCTTAAAGGAGATCG 1511  
QY 1551 TAAGAAACATC-----TTTACGCGGATAACCGTTACGAGGAGAGCCTCT 1595

Db 1512 GATGATTTACGGTGGCCGGTAGAATGTACGGTTCAGGTTGTACGGGAGAACCGTT 1571  
QY 1596 GTTCTCCCGGAGAGAGGA-----GAGGAAGACGAGGATACATCCTCTG 1643  
Db 1572 TTTGGTAGCTAGGATCTCTGTAATCCCGAGCGGAGGAGGATGATGTTATGTTGGTAC 1631  
QY 1644 TTTCTGTCAGGAGGAGACATGGAATCCGAGTTCAGATAGTAAACGCGTTAGCTT 1703  
Db 1632 GTATGTTTACAGTAAGTACCTGGAGATCGAAGTTCTGGTGTAGGACGCTAAATCGCC 1691  
QY 1704 AGAGGTTGAA-----GCAACGGTTAAACTTCCGTCAAGGGTTCGGTACGGAATTCACGG 1757  
Db 1692 GGAGCTTGAATTCGTCCGCGCGGTGAGGTTGCCGGAAGGTTCCGTACGGATCCATGG 1751  
QY 1758 TACATTCATCGGAGCGGATGTTGGCGAAGC 1789  
Db 1752 GTTATTTGTCAAGGAAGTGACCTTAAATAGC 1783

## RESULT 11

US-09-758-269-7  
; Sequence 7, Application US/09758269  
; Patent No. US20020104120A1  
; GENERAL INFORMATION:  
; APPLICANT: IUCHI, SATOSHI  
; APPLICANT: KOBAYASHI, MASATOMO  
; APPLICANT: SHINOZAKI, KAZUO  
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN  
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE  
; FILE REFERENCE: 3914-3  
; CURRENT APPLICATION NUMBER: US/09/758,269  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: JP 2001-003476  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: JP 2000-010056  
; PRIOR FILING DATE: 2000-01-13  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 1617  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1614)  
US-09-758-269-7

Query Match 7.8%; Score 139.8; DB 10; Length 1617;  
Best Local Similarity 49.7%; Pred. No. 1.2e-36;  
Matches 441; Conservative 0; Mismatches 437; Indels 9; Gaps 3;  
QY 392 ATGACAGCCGTCGGCGTAATCTTCGGTGGTCCGGAACCTTCCCGATTCCATCAAG 451  
Db 170 ATGAATCTCTCCCGTCAAGGATCTCCCGTCCATGATTTCTTCCGAAATGTTGATG 229  
QY 452 GAGTGTATGTGCGCAACGAGCTAACCCACTTACAGCGCGGTGACAGGTCAACACTTCT 511  
Db 230 GTGAATTTGTAGGGTTGGTCCAAACCCCAAGTTTGTATGCTGCTGGATATCACTGGT 289  
QY 512 TCGACGGAGAGCGGTATGTTTCAAGCGCTCAATTCGACACGGTTCAGCTAGCTACGCTT 571  
Db 290 TTGATGGAGATGGGATGATTCATGGGTAACGATCAAGATGGGAAGCTACTTATGTTT 349  
QY 572 GCCGGTTTACTCAGACTAACCGGTTTGTTCAGGAACGTCATTTGGTTCGACCGGTTTTC 631  
Db 350 CTCGATATGTTAAGACATCATGCTTAAAGCAGGAAGAGTTCTTCGGAGCTGCCAAATCA 409  
QY 632 CCAAGCCATCGGTAGCTTACGGCCACACCGGTATGCCGCTCATGCTATCTTACG 691  
Db 410 TGAA---GATTGGTGACCTTAAGGGGTTTTTCGATTGCTTAATGGTCAATGTCCACAGC 466  
QY 692 CCAGAGCTGCAGCGGTATAGTCGACCCGCGCACAGGAACCGGTGTAGCTAACCGCGGT 751

Db 467 TGAGACGAGCTCAAAATATTGGACAACACTTATGAAATGAACTGCAATACAGCAC 526  
QY 752 TGGTCTTATTCAATGCGCGTATTGGCTATGTCGGAGGATGATTTTACCTTACCAAGTTC 811  
Db 527 TCGTATATCACCATTGGAACACTTCTAGCATTTACAGGAGCAGATAGCGGTAGCTCATCA 586  
QY 812 AGATCACTCCCAATGGAGATTTAAACACCGTTGGTCCGTTCCGATTTTGTATGGAACAATTAG 871  
Db 587 AAGTTTTGGAAGATGGAGACCTGCAAACTCTTGGTATAATAGATTATGACAAGAGATTGA 646  
QY 872 AATCCACATGATGCGCCACCGGAAAGTCGACCGGAATCCGCGTAACCTTTCGCTTTAA 931  
Db 647 CCACACTCTTCTCATGCTTCAACCAAGTTGACCCGGTTACCGGTGAATGTTTACATTCG 706  
QY 932 GCTACGAGCTGCTTCAAGCCCTTACCTAAATPACTTCCGATTTCTCACCGGACGGAACCTA 991  
Db 707 GCTATTCGCATACGCCA---CCTTATCTCATACAGAGTTATCTCGAAAGATGGCATTA 763  
QY 992 AATCACCAGGCTCGAGATTCAGCTTGTATCAGCAACAGATGATGACGATTTTCGCGATTA 1051  
Db 764 TGCATGACCCAGTCCCAATTAATATATCAGAGCTATCATGATGATGATTTTGTCTATTA 823  
QY 1052 CAGAGAACTTTCGTCTGCTACCTGACCCAGCAAGTCGTTTTTCAAGCTGCCGAGATGATCC 1111  
Db 824 CTGAGACTTATGCAATCTTTCATGATCTTCTATGCACTTCAGGCCAAAGAAATGGTGA 883  
QY 1112 GCGTGGGTCTCCGGT---GGTTTACGACAGACAGGTTCGCAAGATTCGGGATTTTAG 1168  
Db 884 AAGAGAAAGAAATGATATATCTCATTTGATCCCAAAAAGGCTCGTTTTTGGTGTCTTC 943  
QY 1169 ACAAATACGCGCAAGATTCATCGAACATTAAGTGGATTGATGCTCCAGATTTGCTTCTGCT 1228  
Db 944 CACGCTATGCCAAGATGAATTTATGATTAGTTGGTTGAGCTTCCCACTGCTTTATTT 1003  
QY 1229 TCCATCTCTGGAACGCTTGGAGAGCCGCAAGACAGATGAAGTCGTC 1275  
Db 1004 TCCCAACGCCAATGCTTGGGAAGAGAGGATGAAGTCGCTCCTCATC 1050

## RESULT 12

US-09-758-269-17  
; Sequence 17, Application US/09758269  
; Patent No. US20020104120A1  
; GENERAL INFORMATION:  
; APPLICANT: IUCHI, SATOSHI  
; APPLICANT: KOBAYASHI, MASATOMO  
; APPLICANT: SHINOZAKI, KAZUO  
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN  
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE  
; FILE REFERENCE: 3914-3  
; CURRENT APPLICATION NUMBER: US/09/758,269  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: JP 2001-003476  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: JP 2000-010056  
; PRIOR FILING DATE: 2000-01-13  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 17  
; LENGTH: 1617  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1614)  
US-09-758-269-17

Query Match 7.8%; Score 139.8; DB 10; Length 1617;  
Best Local Similarity 49.7%; Pred. No. 1.2e-36;  
Matches 441; Conservative 0; Mismatches 437; Indels 9; Gaps 3;  
QY 392 ATGAACAGCCGTCGGCGTAATCTTCGGTGGTCCGGAACCTTCCCGATTCCATCAAG 451

170 ATGAACTCTCCCGTCAAGGATCTCCCGTCCATGGATTCTTCCGAAATGCTTGAATG 229  
452 GAGTGTATGTGCGCAACGAGCTACCCACTTACAGAGCGGTGACAGGTCAACACTTCT 511  
230 GTGAATTTGTGAGGGTGTGCTCAACACCCCAAGTTTGTGCTGCTGGATATCACTGGT 289  
512 TCGACGAGAGCGGTATGTTTCAACGCGTCAAAATCGAACACGCTTACAGTACGCTT 571  
290 TTGATGGAGATGGATGATTCATGGGTACGCATCAAAAGATGGGAAAGCTACTTATGTTT 349  
572 GCCGGTTTACTCAGACTAACCGGTTTGTTCAGAACGCTCAATTCGGTTCGACCGGTTTCC 631  
350 CTCGATATGTTAAGACATCACTGCTTAAGCAGGAGAGTTCTTTCGGAGCTGCCAAATCA 409  
632 CCAAAGCATCGGTGAGCTTCAACGCCACACCGGTATTCGCCGACTCATGCTATCTACG 691  
410 TGAA---GATTGGTACCTTAAGGGTTTTCGGATTGCTAATGCTCAATATCAACAGC 466  
632 CAGAGCTGCGACCGGTATAGTCAGCCGCGACACGAGAACCGGTGTAGCTTACGCCGTT 751  
467 TGAGAACGAAGCTCAAAATATTGGACAACTTATGGAAATGGAATGCCAATACAGCAC 526  
752 TGGTCTATTCAATGGCGGTTATTGGCTATGTTCGAGGATGATTACCTTACCAAGTTC 811  
527 TCGTATATCACTGGAACCTTCTAGCAATACAGGAGGAGATAGCGGTACGTCACTCA 586  
812 AGATCACTCCCAATGGAGATTAAACCGGTGTCGGTTCGATTTGATGGCAATAG 871  
587 AAGTTTGGAGATGGAGACCTCGCAACTCTTGGTATATAGATTATGACAAGAGATTGA 646  
872 AATCACAATGATTCGCCACCGAAGTCGACCGGAATCGGTGAACCTTTCGCTTAA 931  
647 CCCATCTCTCTGCTCAACCAAGTTGACCGGTTACCGGTGAATGTTTACATTCG 706  
932 GCTACGAGCTCGTTTCAAAGCTTACCTAAATATCTCCGATTCTCAGCGACGAACTA 991  
707 GCTATTGCGATACGCA---CCTTATCTCAGATACAGAGTTATCTCGAAGATGGCATTA 763  
992 AATCAGCGAGCTCGAGATTACGTTGATCAGCAAGCTATGATCAGCATTTCCGATTA 1051  
764 TCGATGACCCGATCCCAATATCTATATCAGAGCCCTATCATGATGATGATTTCGTATTA 823  
1052 CAGAGAACTTCGCTCGTACCTGACCAAGTGTGTTTCAAGTCCGCGAGATGATCC 1111  
824 CTGAGACTTATGCAATCTTCATGATCTTCCTATGCACTTCAGGCCAAAGGAATGGTGA 883  
1112 GCGTGGGTCTCGGT---GGTTTACGACAGAACAGGTCCAGATTCGGATTTAG 1168  
884 AAGAGAAGAAATGATATATCTCACTTGATCCCAAAAAGGCTCGTTTGGTGTCTTC 943  
1169 ACAATAACCGGAAGATTTCATCGAACATTAAGTGGATTGATCTCCAGATTGCTTCTGCT 1228  
944 CGCGCTATGCCAAGGATGAATATGATGATGATGTTTTCAGCTTCCCAACTGCTTATTT 1003  
1229 TCATCTCTGGAACGTTTGGGAGAGCGCAGAACAGATGAAGTCTC 1275  
1004 TCCACAACCCCAATGCTTGGGAGAGAGGATGAAGTCTGCTCTCATC 1050

## RESULT 13

US-09-878-574-2872  
; Sequence 2872, Application US/09878574  
; Patent No. US20020110548A1  
; GENERAL INFORMATION:  
; APPLICANT: Byrum, Joseph R.  
; APPLICANT: Thompson, Michael D.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(15401)B  
; CURRENT APPLICATION NUMBER: US/09/878,574  
; CURRENT FILING DATE: 2001-12-21

; PRIOR APPLICATION NUMBER: 09/333,535  
; PRIOR FILING DATE: 1999-06-14  
; NUMBER OF SEQ ID NOS: 15775  
; SEQ ID NO. 2872  
; LENGTH: 320  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(320)  
; OTHER INFORMATION: unsure at all n locations  
; OTHER INFORMATION: Clone ID: LIB3028-019-Q1-B1-B11  
US-09-878-574-2872

Query Match 5.2%; Score 93; DB 10; Length 320;  
Best Local Similarity 63.5%; Pred. No. 6.4e-21;  
Matches 160; Conservative 0; Mismatches 86; Indels 6; Gaps 1;

QY 1418 AAGTCAACCTCAAGCAGGATGTCACAGAAAACATGTCGCGCTAAACCAAAATCG 1477  
Db 69 AATGAACCTGGAGCAGGATGTTGAACAGGAAAGGTTGGGGAGAAAAACACGGTTG 128  
QY 1478 CTTACTTGGCTTTAGCCGAGCCGTGCTTAAAGTCTCAGGATTCGCTAAAGTTGATCTCA 1537  
Db 129 CATATTGTGCATAGCGGAACCGTGGCCGAGGTTGTCGGGGGTGGCGAAGTGGACCTGG 188  
QY 1538 CTACTCGAGAAGTTAAGAACATCTTTACGCGATAACCGTTACGAGGAGAGCCCTCTCT 1597  
Db 189 AGAGTGGGAAGTTAAGAGGACGAGTCCGAGAGAGAGGTTCCGTGGAGAGCCATTCT 248  
QY 1598 TTCTCCCC-----CGGAGAAGGAGGAGAGAGAGAGGATACATCTCTCTGTTGCTTC 1651  
Db 249 TCTTGCCACACGCTGGAGGAAATGGGAATGAGGATGAGGATGAGGCTTTTGTGC 308  
QY 1652 ACCACGAGA 1663  
Db 309 ATGACGAGATGA 320

## RESULT 14

US-09-770-791-402  
; Sequence 402, Application US/09770791  
; Patent No. US20020062014A1  
; GENERAL INFORMATION:  
; APPLICANT: Gorlach, Jorn  
; APPLICANT: An, Yong-Qiang  
; APPLICANT: Hamilton, Carol M.  
; APPLICANT: Price, Jennifer L.  
; APPLICANT: Raines, Tracy M.  
; APPLICANT: Yu, Yang  
; APPLICANT: Rameaka, Joshua G.  
; APPLICANT: Page, Amy  
; APPLICANT: Matthew, Abraham V.  
; APPLICANT: Ledford, Brooke L.  
; APPLICANT: Woessner, Jeffrey P.  
; APPLICANT: Haas, William David  
; APPLICANT: Garcia, Carlos A.  
; APPLICANT: Krickler, Ted  
; APPLICANT: Slader, Ted  
; APPLICANT: Davis, Keith R.  
; APPLICANT: Allen, Keith  
; APPLICANT: Hoffman, Neil  
; APPLICANT: Hurban, Patrick  
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
; FILE REFERENCE: 2029 (PARA-018PRV)  
; CURRENT APPLICATION NUMBER: US/09/770,791  
; CURRENT FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: 60/178,480  
; PRIOR FILING DATE: 2000-01-27  
; NUMBER OF SEQ ID NOS: 999  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 402



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2003, 15:19:05 ; Search time 117 Seconds  
(without alignments)  
6790.510 Million cell updates/sec

Title: US-09-758-269-5  
Perfect score: 1800  
Sequence: 1 atggcttttttcacggcaac.....tggcgaagcaggtcgtgtga 1800

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	55.6	3.1	4403765	3 US-09-103-840A-2	Sequence 2, Appli
C 2	55.6	3.1	4411529	3 US-09-103-840A-1	Sequence 1, Appli
C 3	53.6	3.0	4403765	3 US-09-103-840A-2	Sequence 2, Appli
C 4	53.6	3.0	4411529	3 US-09-103-840A-1	Sequence 1, Appli
C 5	50	2.8	1518	4 US-08-976-063E-21	Sequence 21, Appli
C 6	50	2.8	32679	4 US-08-976-063E-1	Sequence 1, Appli
C 7	40.4	2.2	1953	4 US-09-252-991A-2481	Sequence 2481, Ap
C 8	38.8	2.2	7218	1 US-08-232-463-14	Sequence 14, Appli
C 9	36.2	2.0	738	4 US-09-252-991A-9920	Sequence 9920, Ap
C 10	36.2	2.0	834	4 US-09-252-991A-10156	Sequence 10156, A
C 11	35.6	2.0	36063	4 US-08-311-731A-140	Sequence 140, App
C 12	34.8	1.9	4695	4 US-09-309-572-9	Sequence 9, Appli
C 13	34.8	1.9	4695	6 5225348	Patent No. 5225348
C 14	34.8	1.9	8450	4 US-09-091-889A-5	Sequence 5, Appli
C 15	34.2	1.9	1221	1 US-08-434-881-1	Sequence 1, Appli
C 16	34.2	1.9	1221	3 US-08-977-771-1	Sequence 1, Appli
C 17	34.2	1.9	1221	3 US-09-361-773-1	Sequence 12, Appli
C 18	34.2	1.9	1370	4 US-09-026-408-12	Sequence 12, Appli
C 19	34.2	1.9	1371	4 US-09-026-408-1	Sequence 1, Appli
C 20	34.2	1.9	1434	4 US-09-016-434-456	Sequence 456, App
C 21	33.6	1.9	456	4 US-09-328-352-3980	Sequence 3980, Ap
C 22	33.4	1.9	396	4 US-09-702-705-21	Sequence 21, Appli
C 23	33.4	1.9	396	4 US-09-736-457-21	Sequence 21, Appli
C 24	33.4	1.9	451	4 US-08-702-705-400	Sequence 400, App
C 25	33.4	1.9	451	4 US-09-736-457-400	Sequence 400, App
C 26	33.4	1.9	454	4 US-09-702-705-1123	Sequence 1123, Ap
C 27	33.4	1.9	454	4 US-09-736-457-1123	Sequence 1123, Ap

C 28	33.4	1.9	457	4 US-09-389-681-297	Sequence 297, App
C 29	33.4	1.9	457	4 US-09-620-405B-297	Sequence 297, App
C 30	33.4	1.9	457	4 US-09-339-338-297	Sequence 297, App
C 31	33.4	1.9	457	4 US-09-433-826B-297	Sequence 297, App
C 32	33.4	1.9	457	4 US-09-604-287A-297	Sequence 297, App
C 33	33.4	1.9	788	3 US-08-991-789A-177	Sequence 177, App
C 34	33.4	1.9	788	4 US-09-062-451-177	Sequence 177, App
C 35	33.4	1.9	788	4 US-09-598-326-177	Sequence 177, App
C 36	33.4	1.9	788	4 US-09-289-198-177	Sequence 177, App
C 37	33.4	1.9	1047	4 US-08-494-907-11	Sequence 11, Appli
C 38	33.4	1.9	1047	5 PCT-US96-10986-11	Sequence 11, Appli
C 39	33.4	1.9	1753	6 5225348-2	Patent No. 5225348
C 40	33.4	1.9	3680	2 US-08-494-907-1	Sequence 1, Appli
C 41	33.4	1.9	3680	5 PCT-US96-10986-1	Sequence 1, Appli
C 42	33.4	1.9	5076	2 US-08-494-907-2	Sequence 2, Appli
C 43	33.4	1.9	5076	5 PCT-US96-10986-2	Sequence 2, Appli
C 44	33.4	1.9	6170	2 US-08-494-907-4	Sequence 4, Appli
C 45	33.4	1.9	6170	5 PCT-US96-10986-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1  
US-09-103-840A-2/c  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2  
Query Match 3.1%; Score 55.6; DB 3; Length 4403765;  
Best Local Similarity 53.7%; Pred. No. 1.6e-05;  
Matches 115; Conservative 0; Mismatches 99; Indels 0; Gaps 0;  
QY 419 CGGTGTCGGAACACTTCCGATTCACATCAAGAGGTATGTGCGCAACGGAGCTAAC 478  
Db 1018507 CCGTGCAGCGGAAGTCCCGCGGACCTGGACGCGATCTACCTGCGCACACCGGAAAC 1018548  
QY 479 CACTTCAGAGCGGTGACAGGTTCACACCTTTCGACGAGAGCGGTATGTTTCACGCCG 538  
Db 1018547 CGCTACACCGCGGATTCGCGACCTACCCCTTCGATGCGGACGATCATGTCG 1018488  
QY 539 TCAATTCGAACACGGTTTCAGCTAGCTTTCGCGGTTTACTAGACTAACCGGTTG 598  
Db 1018487 TCGGCTTCGCTGATGGAAGAGCTTCTACCGAACCGATTTATTCGACCGATGATCT 1018428  
QY 599 TTCAGGACGTCATTTGGTTCGACCGGTTTTCG 632  
Db 1018427 TGGCCGGAACGAGGCGCGGCGGCTGTGGCC 1018394  
RESULT 2  
US-09-103-840A-1/c  
; Sequence 1, Application US/09103840A



```

; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 3.1%; Score 55.6; DB 3; Length 4411529;
Best Local Similarity 53.7%; Pred. No. 1.6e-05;
Matches 115; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 419 CGGTGGTGGAAAATTCCTCCATCCATCAAGAGGTGTATGTGCGCAACGGAGTAACC 478
DB 1018605 CGGTGACAGGGGAGTCCCGCCGACCTGGACGGCATCTACCTGCGCAACACCGAGAACC 1018546

QY 479 CACTTCACGAGCGGTGACAGGTACACCACTTCTTCGACGGAGCGGTATGTTACGCCG 538
DB 1018545 CGCTACACCCGCACTTCGGACCTACCAACCCCTTCGATGGGACGGCATGATCATGTG 1018486

QY 539 TCAATTTGACACACGGTTCAGTACGTACGTTCCGGTTTACTCAGCTAAACGGTTTG 598
DB 1018485 TCGGTTCCGTGATGGAAGAGCTTCTACCGCAACCGATTATTTCGACCGATGATTCT 1018426

QY 599 TTCAGGAACGTCAATTGGGTGACCGGTTTCCC 632
DB 1018425 TGGCCGAGACAGAGCCGCGCGCTGTGCC 1018392

RESULT 3
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 3.0%; Score 53.6; DB 3; Length 4403765;
Best Local Similarity 49.0%; Pred. No. 7.1e-05;
Matches 143; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 820 CCCAATGGAGATTAAACCGTGGTGGTTCGATTTCGATTCGACAAATTAGATCCACA 879

; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529;
Best Local Similarity 49.0%; Pred. No. 7.1e-05;
Matches 143; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 820 CCCAATGGAGATTAAACCGTGGTGGTTCGATTTCGATTCGACAAATTAGATCCACA 879
DB 750412 CTCACCGATGAGCTGGACACCGCTGGGACCTGTGACTTCGACGGACCCCTGCACGCGGT 750471

QY 880 ATGATTGCCACCCGAAAGTCGACCCGGAATCCGGTGAACCTCTTCGCTTTAAGCTACGAC 939
DB 750472 TACACCGCCCATCCGAGCGTGATCCGACACGGGTGAATCGACGCGGTGCTCTACTCG 750531

QY 940 GTCGTTTCAAGCCTTACCTAAATATCTCCGATTCTCACCGGACGGAACTAAATCACCG 999
DB 750532 TTCGCCCGCGGACACAGAGTGCAGTACTCGGTGATCGGACCGACGACACCGCTCGTCG 750591

QY 1000 GACGTGAGATTTCAGCTTCGATTCAGCCCAAGATGATGATGATTCGCGATTACAGAGAC 1059
DB 750592 ACGGTTGATATCGAGGTGGGGATCGCGATGATGATGATGATGATGATGATGATGATG 750651

QY 1060 TTCGTCCTGCTACCTGACGACGAAAGTGGTTTCAAGCTGCCGAGATGATCC 1111
DB 750652 TACGTGGTGTATCTACGACCTGCCGGTGACCTTCGACCCCAATGAGGTGGTGC 750703

RESULT 5
US-08-976-063E-21/c
; Sequence 21, Application US/08976063E
; Patent No. 6524831
; GENERAL INFORMATION:
; APPLICANT: Steinduchel, Alexander

```

APPLICANT: Priefert, Horst  
APPLICANT: Rabenhorst, Jürgen  
TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL  
TITLE OF INVENTION: ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN AND  
TITLE OF INVENTION: VANILLIC ACID AND THEIR USE  
FILE REFERENCE: Beyer-9998-CAO  
CURRENT APPLICATION NUMBER: US/08/976.063E  
CURRENT FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY  
NUMBER OF SEQ ID NOS: 45  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 21  
LENGTH: 1518  
TYPE: DNA  
ORGANISM: not required under old rule  
US-08-976-063E-21

Query Match 2.8%; Score 50; DB 4; Length 1518;  
Best Local Similarity 47.5%; Pred. No. 6.4e-06;  
Matches 149; Conservative 0; Mismatches 165; Indels 0; Gaps 0;  
QY 364 CAGATGCGCGGAAATTTTCTCCGGTGAATGAACAGCGGTCGCGCGTAACTTCCGGTG 423  
DB 1491 CAATTAGTAGGAACACTTCTCCCGACCGGTATAGAGGAGACTTGTTCGATCTAGAGTT 1432  
QY 424 GTGCGAAACTTCCGATTCATCAAGAGGTATGTGCGCAACGGAGCTAACCCACTT 483  
DB 1431 GACGGGAAATCCCAAAATCAATAATGAAGTTCTACCGTAAACGCCAGAGCTCAA 1372  
QY 484 CAGAGCGCGGTGACAGGTGACCACTTCTTCGAGGAGAGCGGTATGGTTACGCGGTCAA 543  
DB 1371 GTTACCCCAAAATTCACACCTTCATAGATGGAGATGGAATGGCTCTGCTTCCAC 1312  
QY 544 TTGCAACAGGTTACGTAGCTAGCTTCCCGGTTTACTCAGCTAACCGGTTTGTTCAG 603  
DB 1311 TTGCAAGATGGTCATGCGACTTATCATGCTGCTGGTTAAACCGCTGATTCAGGCC 1252  
QY 604 GAACGTCAATTGGGTGCGACCGGTTTCCCAAGGCAATCGGTGAGCTTACGCGCACACC 663  
DB 1251 GAACGACTAGCGGAAATCGCTATTGTCATGTACAGAAACCCCTATACCGAGCACCC 1192  
QY 664 GGTATGCCCGACT 677  
DB 1191 AGTGTAAGGACT 1178

RESULT 6  
US-08-976-063E-1/c  
Sequence 1, Application US/08976063E  
Patent No. 6524831  
GENERAL INFORMATION:  
APPLICANT: Steinbuechel, Alexander  
APPLICANT: Priefert, Horst  
APPLICANT: Rabenhorst, Jürgen  
TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL  
TITLE OF INVENTION: ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN AND  
TITLE OF INVENTION: VANILLIC ACID AND THEIR USE  
FILE REFERENCE: Beyer-9998-CAO  
CURRENT APPLICATION NUMBER: US/08/976.063E  
CURRENT FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY  
PRIOR FILING DATE: 1996-11-29  
NUMBER OF SEQ ID NOS: 45  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 32679  
TYPE: DNA  
ORGANISM: Pseudomonas sp.  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (3146)..(3997)  
OTHER INFORMATION: gene = "ORF1"

US-08-976-063E-1

Query Match 2.8%; Score 50; DB 4; Length 32679;  
Best Local Similarity 47.5%; Pred. No. 4.6e-05;  
Matches 149; Conservative 0; Mismatches 165; Indels 0; Gaps 0;  
QY 364 CAGATGCGCGGAAATTTTCTCCGGTGAATGAACAGCGGTCGCGGTAATCTTCCGGTG 423  
DB 16783 CAATTAGTAGGAACACTTCTCCCGACCGGTATAGAGGAGACTTGTTCGATCTAGAGTT 16724  
QY 424 GTGCGAAACTTCCGATTCATCAAGAGGTATGTGCGCAACGGAGCTAACCCACTT 483  
DB 16723 GACGGGAAATCCCAAAATCAATAATGAAGCTTCTACCGTAAATACGCCAGAGCTCAA 16664  
QY 484 CAGAGCGCGGTGACAGGTCAACACTTCTTCGAGGAGAGCGGTATGTTTACGCGCTCAA 543  
DB 16663 GTTACCCCAAAATTCACACCTTCATAGATGGAGATGGAATGGCTCTGCTTCCAC 16604  
QY 544 TTGCAACAGGTTACGTAGCTAGCTTCCCGGTTTACTCAGCTAACCGGTTTGTTCAG 603  
DB 16603 TTGCAAGATGGTCATGCGACTTTCATCATGCTGGGTTTAAACCGCTCGATTCACGGCC 16544  
QY 604 GAACGTCAATTGGGTGCGACCGGTTTCCCAAGGCAATCGGTGAGCTTACGCGCACACC 663  
DB 16543 GAACGACTAGCGGAAATCGCTATTGTCATGTACAGAAACCCCTATACCGAGCACCC 16484  
QY 664 GGTATGCCCGACT 677  
DB 16483 AGTGTAAGGACT 16470

RESULT 7  
US-09-252-991A-2481/c  
Sequence 2481, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252.991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 2481  
LENGTH: 1953  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-2481

Query Match 2.2%; Score 40.4; DB 4; Length 1953;  
Best Local Similarity 44.1%; Pred. No. 0.011;  
Matches 216; Conservative 0; Mismatches 271; Indels 3; Gaps 1;  
QY 273 AGCGCGCGCGGACGCTTGGACGCGGCGGAGGTTTCTTGTGACGCCAGGAGCTACA 332  
DB 640 AGCGAGGATGCGCGACCGCCCGCGGAGGTGGCTTGGAGCATCACCGGGTAGCGA 581  
QY 333 CCGGCTTCTTAAACCGGTGATCTTAGTGTTCAGATCGCGGAAATTTGCTCCGGTGAA 392  
DB 580 TCCGCTCGGCTTCGCGAGGCGCTTCGCGGAGATCGCGGAGCTTGGCTTCGCGCGGG 521  
QY 393 TGAACAGCGCGTCCGGGTGATCTTCGGTGTGCGGAAACTTCCCGATTCATCAAGG 452  
DB 520 TGCAGGCGACCGCGCGGCGATCATGTGCGCGGGCTTCGGTCTTGTTCGCCCATCGGC 461  
QY 453 AGTGTATGTGCGCAACGGAGCTAACCCACTTCA---CGAGCGGTGACAGTCACTT 509  
DB 460 GGATCACTTGCAGCGGCGGCGATGAACCTGATACCGCGCTCGCGCGAGATTTCCGCCA 401

QY 510 CTTGACGAGAGCGGTATGGTTTCAGCGCGTCAAAATTCGAACACAGGTTTCAGCTAGCGC 569  
Db 400 GCTCGGCAATTTCCAGAGAGAGCCATAGCCGGGTGAGGGCTGCCAGCCACTTTCCA 341  
QY 570 TTGCGGTTTACTCAGACTAACCGGTTTGTTCAGAAACGTCGAATGGGTGCGACCGGTTT 629  
Db 340 CGCCAGGTTTCCACAGGCGCGGGTTGAGGTAGCGCGGCGAGCGGTCGCGCCGATGC 281  
QY 630 CCCAAAGCCATCGGTGAGTTTCAGCGCCACACCGGTATTGCCCGACTCATGCTATTCTA 689  
Db 280 TGTGGCTTCGTCGGCGGCTTGACGTGACGCGGTGGGATCGGCTTCGGAGTAGCGG 221  
QY 690 CCCAGAGCTGAGCGCGGTATGTCGACCCCGGCACAGAACCGGTAGCTAACCGCG 749  
Db 220 CCACGAGCGGATGCCATCTCGCGGACGAGCGCGCAGCATCGGACGGCAATCTCGCGC 161  
QY 750 TTTGGTCTAT 759  
Db 160 GGTGGCGAT 151

## RESULT 8

US-08-232-463-14

; Sequence 14, Application US/08232463

; Patent No. 5670367

; GENERAL INFORMATION:

; APPLICANT: DORNER, F.

; APPLICANT: SCHEIFELINGER, F.

; APPLICANT: FALKNER, F. G.

; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

; NUMBER OF SEQUENCES: 52

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley &amp; Lardner

; STREET: 1800 Diagonal Road, Suite 500

; CITY: Alexandria

; STATE: VA

; COUNTRY: USA

; ZIP: 22313-0299

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/232,463

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/97/935,313

; FILING DATE:

; APPLICATION NUMBER: EP 91 114 300.6

; FILING DATE: 26-AUG-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: BENT, Stephen A.

; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 30472/114 IMMU

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703)836-9300

; TELEFAX: (703)683-4109

; TELEX: 899149

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 7218 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; CLONE: pTZgpt-F18

; US-08-232-463-14

Query Match

Best Local Similarity

Matches 16; Conservative 112; Mismatches 74; Indels 0; Gaps 0;

2.2%; Score 38.8; DB 1; Length 7218;

7.9%; Pred.No. 0.087; 74; Indels 0; Gaps 0;

; NUMBER OF SEQ ID NOS: 33142

QY 21 GGCTCGGTTCTCGGAGATGGCTTGGTGGCAATCATCTACGCCCCCATTTATGCTCTTC 80  
Db 1041 GGCTGCGAGGTGAGGAGCTTGGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1100  
QY 81 TCAAAGCTCCGACTTGAGTTATTGTAGCTCTCTTACCTATGCGCAGTGTGTCACAGTAA 140  
Db 1101 YY 1160  
QY 141 GCTCAATGTTTCATCTCGGCTTCACACTCTCCAGTCTCTTATTTCCCTAAGCAATCATC 200  
Db 1161 YY 1220  
QY 201 AAACCTCCCGCCATTGTTCTT 222  
Db 1221 YYYYYYYYYYYYYYYYYYYY 1242

## RESULT 9

US-09-252-991A-9920/c

; Sequence 9920, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 9920

; LENGTH: 738

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-9920

Query Match

Best Local Similarity

Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 393 TGAACAGCCGTCGGCGTAATCTTCGGTGTGGGAAACTTCGGATTCATCAAGG 452  
Db 601 TGAACAGCCGTCGCTGTTGAGCATGTATGGAAGCCAGTTCGCCGCGCCAGG 542  
QY 453 AGTGATGTGCGCAACGAGCTAACCCACTTCACGAGCCGCGTGACAGGTCAACCTTCT 512  
Db 541 AGTCCTGACGCTTCATGCTGCTGAAACCTTGCCCTTGCGGTGCTGGGAGGTCTTCT 482  
QY 513 CGACGAGAGCGGTATGTTCAAGCC 537  
Db 481 CGAAGAGGTGTTGTTAGTTCAAGCC 457

## RESULT 10

US-09-252-991A-10156

; Sequence 10156, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

```
; SEQ ID NO 10156
; LENGTH: 834
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10156

Query Match
Best Local Similarity 2.0%; Score 36.2; DB 4; Length 834;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 393 TGAACAGCGCGTCCGCGGTAATCTTCCGGTGTGCGGAAACTTCCCGGATTCATCAAAAGG 452
Db 249 TGAACAGCGCGTCCGCGGTAATCTTCCGGTGTGCGGAAACTTCCCGGATTCATCAAAAGG 452

QY 453 ACTGATGTGCGCAACGAGCTAACCCACTTACAGAGCGGCTGACAGGTACCACTTCTT 512
Db 309 ACTGCTGCACTTCACTGCTGCTGAAACCTGCGCTTCCGGTGTGCGGAAACTTCCCGG 368

QY 513 CGACGAGAGCGGTATGTTTCAGGCC 537
Db 369 CGAAGAGGTGCTGATGTTTCAGGCC 393

RESULT 11
US-08-311-731A-140
; Sequence 140, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,731A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: C0044/7125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 140:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 36063 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium leprae
US-08-311-731A-140

Query Match
Best Local Similarity 2.0%; Score 35.6; DB 4; Length 36063;
Matches 89; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

; SEQ ID NO 10156
; LENGTH: 834
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10156

Query Match
Best Local Similarity 53.1%; Score 36.2; DB 4; Length 834;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 422 TGTCGGAAGAACTTCCGATTCATCAAAAGGAGTGTATGTGCGCAACGAGGCTAACCCAC 481
Db 22724 TAGCGGAGAAATACCCAGCTACCTAAAAAGCGTACCTACCGAGCGCAACTAAAAACCCAC 22783

QY 482 TTCACGAGCGGTCACAGGTTCACCACTTCTTCGACGAGACGCTATGTTTCACGCCGTCA 541
Db 22784 TGACCCGGAGTTCACAGGTGTATCACCCGTTTCGATGAAGACGCGCTGCTGCACATGGTCG 22843

QY 542 AATTCGAAACACGGTTCAGCTAGCTAGCTGCGGTTTACTCAGATAACCGGTTTGT 599
Db 22844 GCTTCGCGATGAAAAAGCGCTTTTTCGCAACAGCTTTTATAGCACAGACGGATTTTT 22901

RESULT 12
US-09-309-572-9
; Sequence 9, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
; APPLICANT: Heinrich-Pette-Institut
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
; FILE REFERENCE: P50489
; CURRENT APPLICATION NUMBER: US/09/309,572
; CURRENT FILING DATE: 1999-05-11
; EARLIER APPLICATION NUMBER: DE 198 56 463
; EARLIER FILING DATE: 1998-11-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 9
; LENGTH: 4695
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-309-572-9

Query Match
Best Local Similarity 1.9%; Score 34.8; DB 4; Length 4695;
Matches 99; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 340 CCTAAACAGGCTGATCTAGTGTTCAGATGCGCGGAAATTTTCTCCGGTGAATGAACAG 399
Db 3386 CTTGATTGGATTGCCACACGCGCTCATTCGATGCAAGTTTCTGAGCTGAAGGAAAG 3445

QY 400 CCGTCCGCGGTAATCTTCGCGTGTGCGGAAACTTCCCGATTCATCAAGCAGTGTAT 459
Db 3446 ATTGATCCCGTCTCTGTTAAAGCTGGAAGATGGCCCTAAATCTTGAAGTCTGTGTAT 3505

QY 460 GTGCGCAACGAGCTAACCCACTTCACGAGCGGTCAGAGTCCACCTTCTTCGACGGA 519
Db 3506 GTGCCATTGTTGATATGTTCTCTGCGCAAGCCCATGTGTTGAGAGCTTCTCAGACTAT 3565

QY 520 GACGGTATGTTTCACGCCGTCATTT 545
Db 3566 CCACCTTTGGGTAAGGATGACTACTT 3591

RESULT 13
US25348-3
; Patent No. 525348
; APPLICANT: HAGATA, SHIGEKAZU; SUGANO, SUMIO; KIM, DONG W.;
; UETSUKI, TAICHI; KAZIRO, YOSHITO
; TITLE OF INVENTION: DNA FRAGMENT AND EXPRESSION PLASMID
; CONTAINING THE DNA FRAGMENT
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/447,823
; FILING DATE: 08-DEC-1989
; SEQ ID NO:3
; LENGTH: 4695
US25348-3

Query Match
Best Local Similarity 1.9%; Score 34.8; DB 6; Length 4695;
Matches 99; Conservative 0; Mismatches 107; Indels 0; Gaps 0;
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2003, 13:15:00 ; Search time 6756 Seconds  
(without alignments)  
10899.545 Million cell updates/sec

Title: US-09-758-269-5

Perfect score: 1800

Sequence: 1 atggtctcttccacggaac.....tggcgaagcaggtcgtgtga 1800

Scoring table:

IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg.hum.\*

31: em.htg.inv.\*

32: em.htg.other.\*

33: em.htg.mus.\*

34: em.htg.pln.\*

35: em.htg.rtd.\*

36: em.htg.nam.\*

37: em.htg.vrt.\*

38: em.sv.\*

39: em.htgo.hum.\*

40: em.htgo.mus.\*

41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1800	100.0	1800	6	AX148306	AX148306 Sequence
2	1800	100.0	1800	6	BD017431	BD017431 Transgeni
3	1800	100.0	3869	8	AB026549	AB026549 Arabidops
4	1800	100.0	52332	8	AB028617	AB028617 Arabidops
5	1798.4	99.9	2331	8	AY056255	AY056255 Arabidops
6	862.8	47.9	95769	6	AX652128	AX652128 Sequence
7	862.8	47.9	95769	6	AC013430	AC013430 Genomic s
8	824.6	45.8	1818	6	AX148316	AX148316 Sequence
9	824.6	45.8	1818	6	BD017436	BD017436 Transgeni
10	824.6	45.8	2171	8	LENEOXANT	297215 Lycopersico
11	824.6	45.8	19018	8	LESA43079	AJ43079 Lycopersico
12	809.4	45.0	2164	8	STU276244	AJ276244 Solanum t
13	798.4	44.4	1839	6	AX148312	AX148312 Sequence
14	798.4	44.4	1839	6	BD017434	BD017434 Transgeni
15	798.4	44.4	2432	8	AB030293	AB030293 Vigna ung
16	791.6	44.0	2398	8	AF190462	AF190462 Phaseolus
17	750.8	41.7	72058	8	AC074176	AC074176 Arabidops
18	745	41.4	2357	8	AB080192	AB080192 Pisum sat
19	738.2	41.0	1752	6	AX148302	AX148302 Sequence
20	738.2	41.0	1752	6	BD017429	BD017429 Transgeni
21	738.2	41.0	110102	8	ATF28J12	AL021710 Arabidops
22	738.2	41.0	194143	8	ATCHRIV48	AL161548 Arabidops
23	696.8	38.7	2153	8	AF224672	AF224672 Persea am
24	693.8	38.5	2034	8	AB080193	AB080193 Pisum sat
25	665	36.9	2310	8	AF224671	AF224671 Persea am
26	655.8	36.4	175158	2	AC120531	AC120531 Oryza sat
27	655.8	36.4	181159	2	CNS071P7	AL713930 Oryza sat
28	645	35.8	131689	8	CNS08CB9	AL845347 Oryza sat
29	645	35.8	133983	8	CNS08CBP	AL713885 Oryza sat
30	606.6	33.7	1815	6	AX148314	AX148314 Sequence
31	606.6	33.7	1815	6	BD017435	BD017435 Transgeni
32	606.6	33.7	2498	8	ZMU95953	U95953 Zea mays vi
33	573.8	31.9	149097	2	AP005632	AP005632 Oryza sat
34	528.2	29.3	1734	6	AX148310	AX148310 Sequence
35	528.2	29.3	1734	6	BD017433	BD017433 Transgeni
36	528.2	29.3	78921	8	AB028621	AB028621 Arabidops
37	498.4	27.7	82697	8	ATT9A21	AL021713 Arabidops
38	314.6	17.5	602	8	AF399655	AF399655 Lactuca s
39	201	11.2	365	11	BX295824	BX295824 Arabidops
40	195.2	10.8	2016	8	AB080194	AB080194 Pisum sat
41	186.8	10.4	80886	2	AC144759	AC144759 Medicago
42	186.8	10.4	120055	8	AC123596	AC123596 Medicago
43	177.4	9.9	1971	8	MD93765	Z93765 M.domestica
44	176	9.8	120055	8	AC123596	AC123596 Medicago
45	171.2	9.5	1788	8	BT008785	BT008785 Arabidops

# ALIGNMENTS

RESULT 1	AX148306	Sequence 5 from Patent EP1116794.	1800 bp	DNA	linear	PAT 08-JUN-2001
LOCUS	AX148306	Sequence 5 from Patent EP1116794.	1800 bp	DNA	linear	PAT 08-JUN-2001
DEFINITION	AX148306	Sequence 5 from Patent EP1116794.	1800 bp	DNA	linear	PAT 08-JUN-2001
ACCESSION	AX148306	Sequence 5 from Patent EP1116794.	1800 bp	DNA	linear	PAT 08-JUN-2001
VERSION	AX148306.1	GI:14347193	1800 bp	DNA	linear	PAT 08-JUN-2001
KEYWORDS	Arabidopsis thaliana (thale cress)					
SOURCE	Arabidopsis thaliana					
ORGANISM	Arabidopsis thaliana					
REFERENCE	Arabidopsis thaliana					
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
TITLE	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;					
	Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi					
	luchi,S.; Kobayashi,M. and Shinozaki,K.					
	Transgenic plants carrying neoxanthin cleavage enzyme gene					

JOURNAL	Patent: EP 1116794-A 5 18-JUL-2001;
FEATURES	Riken (JP)
source	Location/Qualifiers
	1..1800
	/organism="Arabidopsis thaliana"
	/mol_type="genomic DNA"
	/db_xref="taxon:3702"
CDS	1..1800
	/notes="unamed protein product"
	/codon_start=1
	/protein_id="CAC41194.1"
	/db_xref="GI:14347194"
	/translation="MASFTATAAVSGRWLGNHTOPPLSSOSSDLSSCLPMSARV
	TRKLVSSALHTPALHPFKOSSNSPAIVVKKAKESNTQKNLFORAAALDAEAG
	FLVSHKLPKADPSVOLAGNPAPVNEQVRNLPVVGKLPDSIKGVYVRNGAP
	LHPVTGHPFDCDGMHAKVKEHSASACFTQTNRFVQERQLGRPVFKALGELH
	GHTGIARLMLFYARAAGIVDAHGTGVANAGLVFNGLRLAWSDDLPVQVITPSP
	DLKTVGRFDFDLESTMIAPKVDSPESGELFALSYDVVSKYLPKFRFSPDGTSPD
	VEIQDQPMHDFAITENFVVPDQVFKLPKIRGSPVYDKVNRKVFGLDKY
	AEDSNIKWIDAPDCFHLNWAESPETDEVVIGSCMTPPDSIFNSDENLKSVL
	EIRNLKTEGSTRPFIISNEDQVNLKAEAGVNRNMLGRKTFAYLALAPKPVSGFA
	KVDLTTGEVKHLYGNRYGGPELFPGRGGBEDGYIILCFVHDEKWKSELQIVNAV
	SLEVEATVLPSPVPGFHTFGADDLAKQVV"
BASE COUNT	458 a 464 c 439 g 439 t
ORIGIN	
Query Match	100.0%; Score 1800; DB 6; Length 1800;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1800; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 ATGGCTTCTTACGGCAACGGCTGCGGTTCTCGGAGATGGCTTGGTGGCAATCACT 60
Db	1 ATGGCTTCTTACGGCAACGGCTGCGGTTCTCGGAGATGGCTTGGTGGCAATCACT 60
Qy	61 CAGCGGCATATTGCTTCTCAAGCTCCGACTTCAGTTATTTAGTACTCCTTACCTATG 120
Db	61 CAGCGGCATATTGCTTCTCAAGCTCCGACTTCAGTTATTTAGTACTCCTTACCTATG 120
Qy	121 GCAGTCGTGTACAGTAAGCTCAATGTTTCAATCTCGCGTTCCACACTCTCAGCTCTT 180
Db	121 GCCAGTCGTGTACAGTAAGCTCAATGTTTCAATCTCGCGTTCCACACTCTCAGCTCTT 180
Qy	181 CATTTCCCTAAGCAATCATCAACTCTCCCGCATTTGTTAAAGCCCAAGCAAGAA 240
Db	181 CATTTCCCTAAGCAATCATCAACTCTCCCGCATTTGTTAAAGCCCAAGCAAGAA 240
Qy	241 TCCAACTAAACAGATGAATTTGTTCCAGAGAGCGCGCGCGAGCGTTGGACGGCGG 300
Db	241 TCCAACTAAACAGATGAATTTGTTCCAGAGAGCGCGCGCGAGCGTTGGACGGCGG 300
Qy	301 GAGGTTTCCTGTGACCGAGAGACTACCCGCTTCTTAAACGGTGTATCTAGT 360
Db	301 GAGGTTTCCTGTGACCGAGAGACTACCCGCTTCTTAAACGGTGTATCTAGT 360
Qy	361 GTTCAGATCCCGGAAATTTTGCTCCGGTGAATGAACAGCCGTCGCGGTAATCTCCG 420
Db	361 GTTCAGATCCCGGAAATTTTGCTCCGGTGAATGAACAGCCGTCGCGGTAATCTCCG 420
Qy	421 GTGGTCGGAACACTTCCGATTCATCAAGAGAGTGTATGTCGCAACGGAGTACCCA 480
Db	421 GTGGTCGGAACACTTCCGATTCATCAAGAGAGTGTATGTCGCAACGGAGTACCCA 480
Qy	481 CTTCCAGAGCGGTGACAGGTCAACCACTTCTTCAGCGAGACGGTATGGTTACGCGGTC 540
Db	481 CTTCCAGAGCGGTGACAGGTCAACCACTTCTTCAGCGAGACGGTATGGTTACGCGGTC 540
Qy	541 AAATTCGAACCGGTTACGTAGTACGTTGCGGTTTACTCAGACTAACCGGTTGTT 600
Db	541 AAATTCGAACCGGTTACGTAGTACGTTGCGGTTTACTCAGACTAACCGGTTGTT 600
Qy	601 CAGGAACGTCAATTTGGGTTCGACCGGTTTCCCAAAAGCCATCGGTGAGCTTCAAGCCCA 660
Db	601 CAGGAACGTCAATTTGGGTTCGACCGGTTTCCCAAAAGCCATCGGTGAGCTTCAAGCCCA 660

Qy	661 ACCGGTATTCCCGACTCATCTGCTATTCTAGCCAGAGCTCGACCGGTATAGTCGACCG 720
Db	661 ACCGGTATTCCCGACTCATCTGCTATTCTAGCCAGAGCTCGACCGGTATAGTCGACCG 720
Qy	721 GCACACGAACCGGTGTAGCTAAACGCGGTTTGGTCTATTTCATTAATGCGCGGTTATTGGCT 780
Db	721 GCACACGAACCGGTGTAGCTAAACGCGGTTTGGTCTATTTCATTAATGCGCGGTTATTGGCT 780
Qy	781 ATGTGCGAGGATGATTATTACCTTACCAAGTTTCAGATCACTCCCAATGGAGATTTAAAAC 840
Db	781 ATGTGCGAGGATGATTATTACCTTACCAAGTTTCAGATCACTCCCAATGGAGATTTAAAAC 840
Qy	841 GTTGGTCCGGTTCGATTTTTGTAGTACCAATTAAGATTCACATGATTCGCCACCCGAAAGTC 900
Db	841 GTTGGTCCGGTTCGATTTTTGTAGTACCAATTAAGATTCACATGATTCGCCACCCGAAAGTC 900
Qy	901 GACCCGAATCCGGTGAACTCTTCGCTTTAAGCTAGACGTCGTTTCAAGACCTTACCTA 960
Db	901 GACCCGAATCCGGTGAACTCTTCGCTTTAAGCTAGACGTCGTTTCAAGACCTTACCTA 960
Qy	961 AAATACTTCCGATTCTCACCGGACGGAATTAATTCACCGGACGTCGAGATTCAGCTTGT 1020
Db	961 AAATACTTCCGATTCTCACCGGACGGAATTAATTCACCGGACGTCGAGATTCAGCTTGT 1020
Qy	1021 CAGCCAAACGATGATGACGATTTCCGATTTACAGAGAACTTCGTCGTCGTACCTGACCAG 1080
Db	1021 CAGCCAAACGATGATGACGATTTCCGATTTACAGAGAACTTCGTCGTCGTACCTGACCAG 1080
Qy	1081 CAAGTCGTTTCAAGCTGCGGAGATGATCCGGTGGGTCTCCGGTGGTTTACGACAA 1140
Db	1081 CAAGTCGTTTCAAGCTGCGGAGATGATCCGGTGGGTCTCCGGTGGTTTACGACAA 1140
Qy	1141 AACAAAGTCGCAAGATTCGGGATTTTGAACAATAFACCCGGAAGATTCATCGAATTAAG 1200
Db	1141 AACAAAGTCGCAAGATTCGGGATTTTGAACAATAFACCCGGAAGATTCATCGAATTAAG 1200
Qy	1201 TGGATTGATGCTCCAGATTCGCTTCTGCTTCCATCTCTGGAACGCTTCGGAAGACCCAGAA 1260
Db	1201 TGGATTGATGCTCCAGATTCGCTTCTGCTTCCATCTCTGGAACGCTTCGGAAGACCCAGAA 1260
Qy	1261 ACAGATGAAGTCGCTGATAGGTCCTGTATGATCTCCACAGACTCAATTTTCAACGAG 1320
Db	1261 ACAGATGAAGTCGCTGATAGGTCCTGTATGATCTCCACAGACTCAATTTTCAACGAG 1320
Qy	1321 TCTGACGAGATCTCAAGGTGTCCTGCTGAAATCCGCTGATCTCAAAACCGGTGAA 1380
Db	1321 TCTGACGAGATCTCAAGGTGTCCTGCTGAAATCCGCTGATCTCAAAACCGGTGAA 1380
Qy	1381 TCAACTCGCGCTCGGATCATCTCCACGAAAGATCAACAAGTCAACCTCGAAGCAGGATG 1440
Db	1381 TCAACTCGCGCTCGGATCATCTCCACGAAAGATCAACAAGTCAACCTCGAAGCAGGATG 1440
Qy	1441 GTCAACGAAACATGCTCGCGCTTAAACCAAAATTCGCTTACTTGGCTTTAGCCGAGCG 1500
Db	1441 GTCAACGAAACATGCTCGCGCTTAAACCAAAATTCGCTTACTTGGCTTTAGCCGAGCG 1500
Qy	1501 TGGCTTAAAGTCTCAGGATTCGCTTAAAGTTGATCTCACTACTTGAGAAAGTAAAGAACAT 1560
Db	1501 TGGCTTAAAGTCTCAGGATTCGCTTAAAGTTGATCTCACTACTTGAGAAAGTAAAGAACAT 1560
Qy	1561 CTTTACGCGGATAACCGTTACGAGGAGAGCTCTGTTTCTCCCGGAGAGAGGAGAG 1620
Db	1561 CTTTACGCGGATAACCGTTACGAGGAGAGCTCTGTTTCTCCCGGAGAGAGGAGAG 1620
Qy	1621 GAAGACGAGGATACATCTCTGTTTCTTCCACGACGAGAGACATGGAATTCGAGTTA 1680
Db	1621 GAAGACGAGGATACATCTCTGTTTCTTCCACGACGAGAGACATGGAATTCGAGTTA 1680
Qy	1681 CAGATAGTTAAACCGGTTAGCTTGAAGGAAACCGGTTAACTTCCGTCAGGGTT 1740
Db	1681 CAGATAGTTAAACCGGTTAGCTTGAAGGAAACCGGTTAACTTCCGTCAGGGTT 1740



```

QY 1741 CCGTACGGATTTCACGGTACATTCATCGAGCCGCGATGTTGGCGAACGAGTGGTGTGA 1800
Db 1741 CCGTACGGATTTCACGGTACATTCATCGAGCCGCGATGTTGGCGAACGAGTGGTGTGA 1800

RESULT 2
BD017431 1800 bp DNA linear PAT 27-AUG-2002
LOCUS Transgenic plant using neozanthine cleaving enzyme gene.
DEFINITION
ACCESSION BD017431
VERSION JP 2001258579-A/3.
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
REFERENCE
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1800)
AUTHORS Iuchi, K., Kobayashi, M. and Shinozaki, K.
TITLE Transgenic plant using neozanthine cleaving enzyme gene
JOURNAL Patent: JP 2001258579-A 3 25-SEP-2001;
THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH
COMMENT OS Arabidopsis thaliana (thale cress)
PN JP 2001258579-A/3
PD 25-SEP-2001
PF 11-JAN-2001 JP 2001003476
PI KIYOSHI IUCHI, MASATOMO KOBAYASHI, KAZUO SHINOZAKI PC
C12N15/09,A01H5/00,C12N5/10,C12N9/02,C12N5/00,C12N5/00 CC
Transgenic plant using neozanthine cleaving enzyme gene FH Key
Location/Qualifiers
FT CDS Location/Qualifiers
1..1800
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
BASE COUNT 458 a 464 c 439 g 439 t
ORIGIN
Query Match 100.0%; Score 1800; DB 6; Length 1800;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1800; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTCTTTCACGCCAACGCTGCGGTTCTCGGAGATGCTTGGTGGCAATCATACT 60
Db 1 ATGGCTCTTTCACGCCAACGCTGCGGTTCTCGGAGATGCTTGGTGGCAATCATACT 60

QY 61 CAGCGCCATTATCGCTTCTCAAAGCTCCGACTTGGATTATGAGTCTCCTTACCTATG 120
Db 61 CAGCGCCATTATCGCTTCTCAAAGCTCCGACTTGGATTATGAGTCTCCTTACCTATG 120

QY 121 GCAGTCGTGTACACGTAAGCTCAATGTTTCATCTGCGCTTCACACTCCTCCAGCTTT 180
Db 121 GCAGTCGTGTACACGTAAGCTCAATGTTTCATCTGCGCTTCACACTCCTCCAGCTTT 180

QY 181 CATTTCCCTAAGCAATCATCAAACTCTCCGCGCATTTGTTAAGCCCAAGCCAAAGAA 240
Db 181 CATTTCCCTAAGCAATCATCAAACTCTCCGCGCATTTGTTAAGCCCAAGCCAAAGAA 240

QY 241 TCCAACTAAACAGATGAATTTGTTCCAGAGCGCGCGCGAGCGTTGACGCGCGG 300
Db 241 TCCAACTAAACAGATGAATTTGTTCCAGAGCGCGCGCGAGCGTTGACGCGCGG 300

QY 301 GAGGGTTTCCTTGTACGCCACAGAGAGCTACACCCGCTTCCTAAACCGGCTGATCCTAGT 360
Db 301 GAGGGTTTCCTTGTACGCCACAGAGAGCTACACCCGCTTCCTAAACCGGCTGATCCTAGT 360

QY 361 GTTCAGATCCCGGAATTTTGTCTCGGTGAATGAACACCGCTCGCGGTAACTCTCCG 420
Db 361 GTTCAGATCCCGGAATTTTGTCTCGGTGAATGAACACCGCTCGCGGTAACTCTCCG 420

QY 421 GTGGTCGGAATACTTCCCGATTCCATCAAGAGGTGATGTGCGCAACGAGCTAACCCA 480

```

```

Db 421 GTGGTCGGAATACTTCCCGATTCCATCAAGAGGTGATGTGCGCAACGAGCTAACCCA 480
QY 481 CTTTCAGAGCCGCGTGAACAGGTCACTCTCTTCGACGAGAGCGGTATGTTACGCGCTC 540
Db 481 CTTTCAGAGCCGCGTGAACAGGTCACTCTCTTCGACGAGAGCGGTATGTTACGCGCTC 540
QY 541 AAATTTCGAACACCGGTTTCAGTACGTACGCTTTCGCGGTTTACTCAGACTAACCGGTTTGT 600
Db 541 AAATTTCGAACACCGGTTTCAGTACGTACGCTTTCGCGGTTTACTCAGACTAACCGGTTTGT 600
QY 601 CAGGAACGTCATTAATGGGTGACCGGTTTCCCAAGACCATCGGTAGCTTACCGGCCAC 660
Db 601 CAGGAACGTCATTAATGGGTGACCGGTTTCCCAAGACCATCGGTAGCTTACCGGCCAC 660
QY 661 ACCGGTATTTCGCGGACTCATGTATTCTACGCCAGAGCTGCAGCGGTATAGTCGACCCG 720
Db 661 ACCGGTATTTCGCGGACTCATGTATTCTACGCCAGAGCTGCAGCGGTATAGTCGACCCG 720
QY 721 GCACCGGAACCGGTTAGCTAACCGGTTTGGTCTATTTCATAGCGCGGTTATTGGCT 780
Db 721 GCACCGGAACCGGTTAGCTAACCGGTTTGGTCTATTTCATAGCGCGGTTATTGGCT 780
QY 781 ATGTTCGGAGGATGATTTTACCTTACCAAGTTTCAGATCACTCCCAATGGAGATTAAAAAC 840
Db 781 ATGTTCGGAGGATGATTTTACCTTACCAAGTTTCAGATCACTCCCAATGGAGATTAAAAAC 840
QY 841 GTTGGTTCGGTTCGATTTCGATGACAAATTAGAATCCAAATGATTCGCCCAACCGGAAAGTC 900
Db 841 GTTGGTTCGGTTCGATTTCGATGACAAATTAGAATCCAAATGATTCGCCCAACCGGAAAGTC 900
QY 901 GACCCGGAATCCGCTGACTCTTCGCTTTAAGTACGAGCTCGTTTCAAGCTTACCTTACCTA 960
Db 901 GACCCGGAATCCGCTGACTCTTCGCTTTAAGTACGAGCTCGTTTCAAGCTTACCTTACCTA 960
QY 961 AAATCTTTCGGATTCTCACCGGACGGAATTAATTCACCGGACGCTCGAGATTTCAGTTGAT 1020
Db 961 AAATCTTTCGGATTCTCACCGGACGGAATTAATTCACCGGACGCTCGAGATTTCAGTTGAT 1020
QY 1021 CAGCAACGATGATGACGATTTTCGCGATTACAGAGACTTCGCTCGTACCTGACCCAG 1080
Db 1021 CAGCAACGATGATGACGATTTTCGCGATTACAGAGACTTCGCTCGTACCTGACCCAG 1080
QY 1081 CAAGTCGTTTTCAAGCTGCGGAGATGATCCGCGGTGGGTCTCCGTTGGTTTACGCAAG 1140
Db 1081 CAAGTCGTTTTCAAGCTGCGGAGATGATCCGCGGTGGGTCTCCGTTGGTTTACGCAAG 1140
QY 1141 AACAGTTCGCAAGATTCGGGATTTTAGACAAATACGCGAGATTCATCGAACATTAAG 1200
Db 1141 AACAGTTCGCAAGATTCGGGATTTTAGACAAATACGCGAGATTCATCGAACATTAAG 1200
QY 1201 TGGATTGATGCTCCAGATTGCTTCTGCTTCCATCTCTGGAACGCTTGGGAAGACCGAGAA 1260
Db 1201 TGGATTGATGCTCCAGATTGCTTCTGCTTCCATCTCTGGAACGCTTGGGAAGACCGAGAA 1260
QY 1261 ACAGATGAAGTCGTTGATGAGGTCTGTAACCTCCACAGACTCAATTTTCAACGAG 1320
Db 1261 ACAGATGAAGTCGTTGATGAGGTCTGTAACCTCCACAGACTCAATTTTCAACGAG 1320
QY 1321 TCTGACGAGATCTCAAGAGTGTCTGTAATCGGCTGAATCTCAAAACCGGTGAA 1380
Db 1321 TCTGACGAGATCTCAAGAGTGTCTGTAATCGGCTGAATCTCAAAACCGGTGAA 1380
QY 1381 TCAACTCCGCGTCCGATCATCTCCACGAGAGATCAACAAAGTCAACCTCGAAGCAGGATG 1440
Db 1381 TCAACTCCGCGTCCGATCATCTCCACGAGAGATCAACAAAGTCAACCTCGAAGCAGGATG 1440
QY 1441 GTCAACAGAAACATGTCGCGCGTAAACCAAAATTCGTTACTTTCGCTTTAGCGCGCG 1500
Db 1441 GTCAACAGAAACATGTCGCGCGTAAACCAAAATTCGTTACTTTCGCTTTAGCGCGCG 1500
QY 1501 TGGCCTAAAGTCTCAGGATTTCGCTTAAAGTTGATCTCACTACTCGAGAGATTAAAGAACAT 1560
Db 1501 TGGCCTAAAGTCTCAGGATTTCGCTTAAAGTTGATCTCACTACTCGAGAGATTAAAGAACAT 1560

```





```
31021. .31133,31239. .31298,31376. .31492,31588. .31649,
31733. .31793)
/notes="contains similarity to RNA-binding protein
gene_id:MOA2.5"
/codon_start=1
/evidence=not experimental
/db_xref="GI:11994215"
/translation="MATVTEMPDVAVVDDVKEVSKSEKIIDEGIKSIDSXTET
ESLDHKLVAPEKKNPLAKERFPYYPDPKKNQVAKANQLPADDPTTKQSGEE
FDLDXKDDORAHREVRVHMLTGIDKAFQADIDLOTEALRRVVAAGGLGG
FNSQGVQDCRICDPHSLVRFAPFADQQAHEALSGGVLGVPVRLVPSKTAI
LPVNPFLPSREBEMCTRIYCTINDKVSQADVNFESACGEVTRLLGGDLH
STRIAFEPALADSALSNCSGMVWGSPQIRVSPSKTPVRPRTPPSTN"
complement(32182..36456)
/notes="gene_id:MOA2.6"
/codon_start=1
/evidence=not experimental
/product="disease resistance complex protein"
/protein_id="BAB01338.1"
/db_xref="GI:11994216"
/translation="MANSYSSCANVMVERINTSQBELVELCKGSSALLKRLKVALV
TANPLADARORAHREVRVHMLTGIDKAFQADIDLOTEALRRVVAAGGLGG
LPQNLAGRAIRKQKIBPKMEKVRLLEHVHLEVLGKEYSETREPQWASRRP
DDLPQGLGRVEDKALNLLSDDEISIGKPAVSVQMPGVGKTLTEIVENDYR
VTEHFEYKMSIAGINFTVTKAVLQDITSSAVNTEDLPSQIKKILSGKRELL
VLDPFSESSEMSFQVPTDAEBSKIVLTRSEIIVTAKAEKIYKQKMLNBECC
WELTRFAGNIVSGVINOELGIGKIAQCKGLPLAARASHLRSPNDWYAV
SKNFSSYNTGLVPLKLSYDPLQKRCFALGIFPKGHVFDEELVLLMAIDLLY
OPRRSLEDIGNDLGDLVAQFPORLDTWTSFVNHDLMDLAKAVSGDFCRLED
DNIFEIPTSTHRSFSDQDASVAFSICGABFRTILPNSFTLSLESLOLTKVLN
PLNALSGRLISLSHQVITNPKSLGKLKLYLDLSSTKIKELPEFVTCNLQTL
LLSNCRLTSLPGLSIAELINLRLDLVPLVEMPPGIRKLRSLQKLSNFIQRLSGA
GLHELKSLHRLGTRIRSELONAFASEADAGLRKPPFLDGLILKMWKVGSPVGS
TANPLADARORAHREVRVHMLTGIDKAFQADIDLOTEALRRVVAAGGLGG
FNSQGVQDCRICDPHSLVRFAPFADQQAHEALSGGVLGVPVRLVPSKTAI
LPVNPFLPSREBEMCTRIYCTINDKVSQADVNFESACGEVTRLLGGDLH
STRIAFEPALADSALSNCSGMVWGSPQIRVSPSKTPVRPRTPPSTN"
complement(32182..36456)
/notes="gene_id:MOA2.7"
/codon_start=1
/evidence=not experimental
/product="disease resistance complex protein"
/protein_id="BAB01339.1"
/db_xref="GI:11994217"
/translation="MTGIGEMFLAFLQALPQLVSBPSPFRRKRELNENLRLST
ALLTITAVLIDAEKQITNPVEKWNELROVVHAEDALDIAETALRLNIGRESS
SNRLQLRGRSLDFLDGNSSEHLETRLEKVTIRLERASQNRNLGUKELTAMIPKOR
LPTSLVDESEVFGDDDKDIMEIRFLIPENGKONGITVAIVGIGVGKLTLSOLLN
DCHVSYFGTKWAVHSEDFVFKITKVAQVSVSPCEFTDLVLOVKLRLTG
LPFLVLDLWNEFADMLLROPTTAAQSQILVTRSORVASIWCVAHVHNLQPL
SGDCWSLPMKTVFCNQBPCLNREIGDLAERIVHKRGLPLAVKTLGVLRFEGKVE
WKLVSRRINDPLADKSNLLPVLRVSYVYLPAHLKRCFAYCSIIPKHAPEKQKVL
WNAEFLQOTRSKLEBELNGEYSELESLLQTKTRYIMHDFINELAQFASGETS
SKFEDCKLQWSETRYLSYNDVNAEPMFEAREVFLRTFLPLSLNLEKSPSCCLD
QWVSEKLLPTLRVLSLRYKIARLPDPFKNISHAFLDLRLTELEKPLKSLCY
VNLQTLILSYSSKLELPTDINLRYLDLIGTKLRQMPRRFKRLREIDFVMTGTS
SADGSRISLGGDLHGLKIVELQVADVADAAEANLNSKKHLREIDFVMTGTS
SENNTNPHRTONEAREVTEKPLRPHIEKLAIERYKGRFPDMLSDPFRSVIRLR
EQCYCTSLPSLQCLKELHISGMVQSLGKEVFSQDLRQDQOQPPRSLTFR
DNLPMQEWLVRVTRGDLFPLSKLFIIRCEPLTGLTLPPLSLIHLIVKGLLP
QPDHHEYSYRNLQTLISKSCDILVKFLPLNHPANLDKLEVDQCTSLYSELSNHLRG
PNALNLRINDCQNLQPLKALNLPQLVITTCRYLRQPMQEQPQVHHPFHEHUPRS
NVSGSPKSHGSHRSYDSRSSRYD"
42345. .42872
```

CDS

CDS

CDS

CDS

```
/notes="gene_id:MOA2.8
unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAB01340.1"
/db_xref="GI:11994218"
/translation="MSDLVIALVAFVFLVFWFILLSCIMGEXDVHSPPLPRPGOLR
MPVYTRKDLGGDTAISVAGTDIYGGSGVGGHGGRIITSDTKLPSPPSPRLP
PRVQRPKSVSRTKRELIDITGATMLVASSLLTSGSSASSCGSGSHGGCGGGGGG
GGLGGCGGGGGGG"
complement(join(42980..43665,43561..43668,43952..44119,
44215..44460,44539..44677,44759..44977,45060..45754))
/notes="gene_id:MOA2.9"
/codon_start=1
/evidence=not experimental
/product="terpene synthase-like protein"
/protein_id="BAB01341.1"
/db_xref="GI:11994219"
/translation="MKPKVRMLKSSKSNDRIRILHLLNMGIAVHFEIDEILG
QAFNLDDITAKENDLEITSTMEVPRLRGVMPCKSIGDLSHICIHYNLAFDFT
VDANRFGEDGRFKESLAEDIRGMLQLYEAHLGTPSEIDMEALSFTYRLLESFTS
NHTAPSHUSKHLONALYRARYHNLEILVAREYISYEQEDHDETLAKFALNPNY
COLHYIDELKDLTKWKBELDLASKLPVIRDI VEYVFGALALIFEPYISLGRILVTKI
TMIVTFNDTCDAYTLPESVLSVDFQRMDLGDIEKLPSTKIVPRGVETLEISQ
EMRPGASRIQVAVDEIKKLKAYLAISKMARASHVPTFEYMEFGQSMDFHAY
SFIAEDENOTCEWYKSRPKMMEALGVFRINKDINTFEQMSRGEVAKLNCYMK
OHGYSKEAIGEMNKYISNYKILIMEVYLTAVPRPILFTFDMLEKPLIKITLVM
FRLVNDVGTVEHLKLTETSRGVAAGLDFYMKQGVTKKEASOELRKMKNKVV
MEEFMNHDLHPQVFRCHNIAIFDVFTVEVDGYCDPKGIENPWSLYLHPINTP
Query Match 100.0%; Score 1800; DB 9; Length 52232;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1800; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCTTTTTCACGGCAACGGCTGCGGTTCTTGGGAGATGCTTGTGGCAATCATACT 60
DB 13669 ATGGCTTTTTCACGGCAACGGCTGCGGTTCTTGGGAGATGCTTGTGGCAATCATACT 13610
QY 61 CAGCGCCCATATTCGCTTCTTCAAGCTCCGACTTGGATTATGTTAGCTCTTACCTATG 120
DB 13609 CAGCGCCCATATTCGCTTCTTCAAGCTCCGACTTGGATTATGTTAGCTCTTACCTATG 13550
QY 121 GCAGTCGTGTACAGTAAGTCAATGTTTCATCGCTTCCACACCTCCAGCTCTT 180
DB 13549 GCAGTCGTGTACAGTAAGTCAATGTTTCATCGCTTCCACACCTCCAGCTCTT 13490
QY 181 CATTTCCCTTAAGCAATCATCAAACTCTCCCGCATTTGTTTAAAGCCCAAGCCAAAGAA 240
DB 13489 CATTTCCCTTAAGCAATCATCAAACTCTCCCGCATTTGTTTAAAGCCCAAGCCAAAGAA 13430
QY 241 TCCAACTAAACAGATGAATTTGTTCCAGAGAGCGGGCGGCGGAGCTTGACCGGGC 300
DB 13429 TCCAACTAAACAGATGAATTTGTTCCAGAGAGCGGGCGGCGGAGCTTGACCGGGC 13370
QY 301 GAGGGTTTCTTGTGAGCCACAGAGAGCTACACCGGCTTCTTAAACCGCTGATCCTAGT 360
DB 13369 GAGGGTTTCTTGTGAGCCACAGAGAGCTACACCGGCTTCTTAAACCGCTGATCCTAGT 13310
QY 361 GTTCAGATCCCGGAAATTTGCTCGGTGAATGAACAGCCCGCTCGGCGTAACTCTTCG 420
DB 13309 GTTCAGATCCCGGAAATTTGCTCGGTGAATGAACAGCCCGCTCGGCGTAACTCTTCG 13250
QY 421 GTCGTCGAAAACCTCCCGATCCATCAAGAGTGATGTGTGCGCAACGAGAGCTAACCCA 480
DB 13249 GTCGTCGAAAACCTCCCGATCCATCAAGAGTGATGTGTGCGCAACGAGAGCTAACCCA 13190
QY 481 CTTTCAGAGCGGTGACAGGTACACCTTCTTCGAGCGAGAGCGTATGTTTACGCGCTC 540
DB 13189 CTTTCAGAGCGGTGACAGGTACACCTTCTTCGAGCGAGAGCGTATGTTTACGCGCTC 13130
QY 541 AAATTGACACACGGTTACGTAGTACGTTGCGGTTTACTCAGACTAACCGGTTTGT 600
DB 13129 AAATTGACACACGGTTACGTAGTACGTTGCGGTTTACTCAGACTAACCGGTTTGT 13070
```



```
(FLC-1) as a BamHI/XhoI insert.
ecotype: Columbia
1. .2331
/gene="At3g14440"
1. .122
/gene="At3g14440"
123. .1922
/gene="At3g14440"
123. .1922
/gene="At3g14440"
/codon_start=1
/evidence=experimental
/product="putative 9-cis-epoxycarotenoid dioxygenase"
/protein_id="AAL07104.1"
/db_xref="GI:15810433"
/translation="MASTATAAASVRLGNGHTOPPLSSQSDLSYCSLPMASRV
FLVSHKLPKLPKADPSVQIAGNAPNEQVRENLPVVKLPDSIKGVVRNGANP
TRKLVSSALHPKLPKADPSVQIAGNAPNEQVRENLPVVKLPDSIKGVVRNGANP
LHEVPTGHFFDGDGMVHAKVHESASVACQTQINRFVQERQGRVFPKAGELH
GHTGARLMLFYARAAGIYDPAHGTGVANAGLVFNGLRLAMSDDLPYQVQITPND
DLKTVGRFDFDGLSTMIAPKVPDPESGELFALSYDVVSKPLKYFFSPDGTSPD
VLELDQPTMMDFAITENFVVVPOQVVKFLPEMLIGSPVVDKVNKVFARGLDKY
AEDSNIKWIDAPDCFCFLMNAWEPETDEVVWIGSCWTPDPSIFNEDSNILKSVLS
ERLNKLGESTRRPILSNEDQVNLNLEAGVNRNMLGRKTFAYIALAEPWPVKVGFPA
KVDLITGEVKNKHLGDNRYGGEFLFLPGEGBEDSEGLILCFVHDEKTKWSELIQVNAV
SLEVEATVKLPGRVPYGFPHGTFIGADDLAKQVV"
misc_difference 1234
/gene="At3g14440"
/notes="compared to genomic sequence resulting in an amino
acid sequence difference"
/replaces="g"
1923. .2331
/gene="At3g14440"
2059
/misc_difference 2059
/gene="At3g14440"
/notes="not present in genomic sequence"
2312
/misc_difference 2312
/gene="At3g14440"
/notes="not present in genomic sequence"
2315
/misc_difference 2315
/gene="At3g14440"
/notes="not present in genomic sequence"
615 a 560 c 533 g 623 t
BASE COUNT
ORIGIN
Query Match 99.9%; Score 1798.4; DB 8; Length 2331;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1799; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGGCTTCTTTACGGCAACCGGTCGGGTTTCTGGAGATGGCTTGTGGCAATCATCT 60
DB 123 ATGGCTTCTTTACGGCAACCGGTCGGGTTTCTGGAGATGGCTTGTGGCAATCATCT 182
QY 61 CAGCGGCATTATGCTTCTCAAGCTCCGACTGGATTATGTAGCTCCTTACCTATG 120
DB 183 CAGCGGCATTATGCTTCTCAAGCTCCGACTGGATTATGTAGCTCCTTACCTATG 242
QY 121 GCCAGTGTGTACACGTAAGCTCAATGTTTCATCTGGCTTTCACACTCTCCAGCTCTT 180
DB 243 GCCAGTGTGTACACGTAAGCTCAATGTTTCATCTGGCTTTCACACTCTCCAGCTCTT 302
QY 181 CATTTCCCTAAGCAATCATCAAACTCTCCGCCATTGTTGTAGCCCAAGCCAAAGAA 240
DB 303 CATTTCCCTAAGCAATCATCAAACTCTCCGCCATTGTTGTAGCCCAAGCCAAAGAA 362
QY 241 TCCCAACACTAAACAGATGAATTTGTTCCAGAGAGCGCGCGGCGAGCGTTGGACGCGCG 300
DB 363 TCCCAACACTAAACAGATGAATTTGTTCCAGAGAGCGCGCGGCGAGCGTTGGACGCGCG 422
QY 301 GAGGTTTCTTGTGAGCCACAGAGAGTACACCGCTTCTTAAACGGCTGATCTTAGT 360
DB 423 GAGGTTTCTTGTGAGCCACAGAGAGTACACCGCTTCTTAAACGGCTGATCTTAGT 482
QY 361 GTTCAGATCGCGGAATTTTCTCGGTGAATGAACAGCGCGTCCGCGGTAATCTTCGG 420
```



Qy	1501	TGCGCTAAAGTCTCAGGATTCGCTAAAGTTGATCTCACTACTGGAGAAAGTTAAGAAACAT	1560
Db	1623	TGCGCTAAAGTCTCAGGATTCGCTAAAGTTGATCTCACTACTGGAGAAAGTTAAGAAACAT	1682
Qy	1561	CTTTACCGCGGATAACCGTTACGAGGAGAGCCCTCTGTTTCTCCCGGAGAAGGAGGAGAG	1620
Db	1683	CTTTACCGCGGATAACCGTTACGAGGAGAGCCCTCTGTTTCTCCCGGAGAAGGAGGAGAG	1742
Qy	1621	GAAGACGAAGGATACATCTCTGTTTTCGTTTCAGACGAGAGAGACATGGAATTCGAGTTA	1680
Db	1743	GAAGACGAAGGATACATCTCTGTTTTCGTTTCAGACGAGAGAGACATGGAATTCGAGTTA	1802
Qy	1681	CAGATAGTTTAAACCGCGTTAGCTTAGAGGTTTGAAGCAACGGTTTAAACTTCCGTCGAAGGTT	1740
Db	1803	CAGATAGTTTAAACCGCGTTAGCTTAGAGGTTTGAAGCAACGGTTTAAACTTCCGTCGAAGGTT	1862
Qy	1741	CGGTACGGATTTACCGGTACATTCATCGAGCGCGATGATTTGGCGAACACAGGTCTGTGTA	1800
Db	1863	CGGTACGGATTTACCGGTACATTCATCGAGCGCGATGATTTGGCGAACACAGGTCTGTGTA	1922
RESULT 6			
LOCUS		AX652128	95769 bp DNA linear PAT 22-MAR-2003
DEFINITION		Sequence 1979 from Patent WO03000898.	
ACCESSION		AX652128	
VERSION		AX652128.1 GI:29154942	
KEYWORDS		Arabidopsis thaliana (thale cress)	
SOURCE		Arabidopsis thaliana	
ORGANISM		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi	
REFERENCE			
AUTHORS		Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M., Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.	
TITLE		Plant genes involved in defense against pathogens	
JOURNAL		Patent: WO 03000898-A 1979 03-JAN-2003;	
FEATURES		Syngenta Participations AG (CH)	
source			
BASE COUNT		Location/Qualifiers	
ORIGIN		1..95769	
		/organism="Arabidopsis thaliana"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:3702"	
		30958 a 16961 c 16682 g 31168 t	
Query Match			
Best Local Similarity		47.9%; Score 862.8; DB 6; Length 95769;	
Matches 1139; Conservative 0; Mismatches 407; Indels 12; Gaps 2;			
Qy	245	ACACTAAACAGATGAATTTGTTCCACAGAGCGCGCGACGCTTGGACGCGCGGAGG	304
Db	33987	AGATCCAAATTAACCCGGTCAGAGAACCGCGCTATGTTTACGCGGTGGAA	34046
Qy	305	GTTTCCTGTGACCCACGAG---AAGCTACACCGCTTCTTAAACCGCTGATCTAGTG	361
Db	34047	ACGCCATGATCTCACACGAGCGCGCGCTATCCCTATCTTAAACCGCGGATCTCGCG	34106
Qy	362	TTCAGATCGCCGGAATTTTTCCTCCGTTGAATGACAGCCCGTCGCGGTATCTTCGG	421
Db	34107	TCCAAATAGCCGGGAACCTTCTCCCGGTACCGGAGAACCTGCTGTCATACCTTCCGG	34166
Qy	422	TGTCGGAACACTTCCCGATTCATCAAGAGGATGTATGTGCGCAACGGAGCTAACCCAC	481
Db	34167	TGACTGGACAGTACCAAGATGCTTCAAGGAGTTTACGTCAGAACCGAGCAATCCAC	34226
Qy	482	TTACGAGCGGTGACAGGTACCACTTCTTCGCGGAGACGGTATGTTTACGCGGCA	541
Db	34227	TTCAAAACAGTCTCCGGCACCATTTGTTTCGCGGAGACGGTATGTTTACAGTGTCC	34286
Qy	542	AATTTCGAACACGGTTACGTAGCTAGCTTTCGGGTTTACTCAGACTAACCGGTTTGTTC	601



```

Db      35358 AAGACGAGGTTACATAATGTTGTTCTTACAGCAGGAGGAGGTGAATCGGAACGCC 35417
LOCUS   1682 AGATAGTTAACGCGGTTAGCTTAGAGGTTGAACCAACGGTTAAACTTCGTCGAAGGGTTC 1741
        |||||
Db      35418 AACTCATAACGCGGTTAATATGAAGCTTGAAGCTTACCGTACGCTTCGTCGAGAGTGC 35477
        |||||
Qy      1742 CGTACGATTACGGTACATTCATCGAGCCCATGATTTGGCAGACGAGTCGTGTG 1799
        |||||
Db      35478 CTTACGCTTCCACGGAACGCTTCATCAGTAAGGAAGATTATCGAAGCAAGCTTTGTG 35535
        |||||

RESULT 7
ACOL13430
DEFINITION Genomic sequence for Arabidopsis thaliana BAC F3f9 from chromosome
            I, complete sequence.
ACCESSION ACOL13430
VERSION   ACOL13430.5 GI:8096768
KEYWORDS  HTG.
SOURCE    Arabidopsis thaliana (thale cress)
ORGANISM  Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 95769)
AUTHORS   Chao,Q., Brooks,S., Buehler,E., Johnson-Hopson,C., Khan,S., Kim,C.,
            Shinn,P., Altafi,H., Bei,Q., Chin,C., Chlou,J., Choi,E., Conn,L.,
            Conway,A., Gonzales,A., Hansen,N., Howing,B., Koo,T., Lam,B.,
            Lee,J., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S., Liu,A., Liu,J.,
            Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A.,
            Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Federspiel,N.A.,
            Theologis,A. and Ecker,J.R.
            Genomic sequence for Arabidopsis thaliana BAC F3f9 from chromosome
            I
TITLE      Unpublished
JOURNAL    2 (bases 1 to 95769)
REFERENCE  Ecker,J.R.
AUTHORS
TITLE      Direct Submission
JOURNAL    Submitted (11-NOV-1999) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE  3 (bases 1 to 95769)
AUTHORS   Ecker,J.R.
TITLE      Direct Submission
JOURNAL    Submitted (07-FEB-2000) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE  4 (bases 1 to 95769)
AUTHORS   Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q.,
            Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C.,
            Chlou,J., Choi,E., Conn,L., Conway,A., Gonzales,A., Hansen,N.,
            Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J.,
            Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H.,
            Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M.,
            Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
            Direct Submission
JOURNAL    Submitted (24-MAY-2000) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th and
            Hamilton Walk, Philadelphia, PA 19104-6018, USA
REFERENCE  5 (bases 1 to 95769)
AUTHORS   Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q.,
            Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C.,
            Chlou,J., Choi,E., Conn,L., Conway,A., Gonzales,A., Hansen,N.,
            Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J.,
            Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H.,
            Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M.,
            Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
            Direct Submission
JOURNAL    Submitted (26-MAY-2000) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th and
            Hamilton Walk, Philadelphia, PA 19104-6018, USA
COMMENT   On May 26, 2000 this sequence version replaced gi:6921155.
FEATURES   Location/Qualifiers

```

```

1. 95769
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F3f9"
join(71..402,574..649,733..933)
/notes="similar to oxidosqualene cyclase dbj|BAA33462.1"
/codon_start=1
/evidence=not_experimental
/product="F3f9.1"
/protein_id="AAF71792.1"
/db_xref="GI:8052528"
/translation="MGHVQLTFCYFFSFFPFFSFLFLFFLNKXNVILLISGDLTVGVV
FELEAKVEQVPPVKVEDGSTTHETRRMYGEFTSSRLCRPMVTGLSKSLDLSSS
ELRWYVSSFLGLEETFDAREHEELRYVRLNDDGGGLHVEGKSFMECTALNYI
CLRLREGDEGREYACKLTKMPWVTFVLLIYNNSCFPSI"
complement(join(3438..3746,3868..4032,4110..4664))
/notes="similar to F-box domain gb|AAF14684.1"
/codon_start=1
/evidence=not_experimental
/product="F3f9.2"
/protein_id="AAF71813.1"
/db_xref="GI:8052549"
/translation="MLASWQLSTSMKTKNLLILANGRFAPTEGDNVILKLPSTEN
HPSRTLSLYIYILQDAFNCKNLYLDLQFLFAVDVFNVEISSCPSLEVLILQI
IFFNPSEHLKTDHKLRLILSMCNQIDSIIEVRAARDLILSIEYIPCSDDNVLEPRL
QFNGYVAGRLPHTSINISCPQKQESNGMWKTDYATSPASLSLDLKHPRV
EVLAKILAVTEKMEIVEIGFKNNPGLGEGNSDGGQNNLMEKAPFPNADPRVDT
IWMNFKGSNKEQFALASRFVMQKTVMKWKWIKTSFDEKKKETEAAVAKLKLPLKGN
BELSIECF"
join(7138..7400,7491..7887)
/notes="unknown protein; similar to EST gb|AI999747.1"
/codon_start=1
/evidence=not_experimental
/product="F3f9.3"
/protein_id="AAF71793.1"
/db_xref="GI:8052529"
/translation="MERGLNLLKLSLVVVVGSYACAPAPWPNPSNGRPGTCDHVE
CPTYKLVDAGYGEIRFMDALWISTFISLSMTQATKTFRPLNRYIEGDNKSVK
NMTPAVTAQATPGRSVTVSLYLPKXNQNPPOADDLHVRSRTPTVAVRQIGGVYS
NNVAKDEAALMESLRDSNWILPIEKSGKGLPAYFLAVYINPPSHTTARVINEIWPEN
M"
join(8524..8759,8846..9287)
/notes="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F3f9.4"
/protein_id="AAF71794.1"
/db_xref="GI:8052530"
/translation="MEALSIKLSFLITLLSGGSLDLSLPPNCRNACPSYEVVHA
GNGYELHYNTVMTSTPIODISLNEASGNGWQLSDYNGNNDYHORIATLPYIT
QVNSLSTFVSEFVPKAFQBDPPGNLHVQRWDSRYAVKQISGVADHKIGQVA
ELKASLGTVWAKALEKSRFEGVGSWAYVVAQSFQWSQVNEIWFPEMEDEE
TVSIVTQ"
join(13444..13814,13956..14310,14399..14662)
/notes="similar to dioxygenase protein"
/codon_start=1
/evidence=not_experimental
/product="F3f9.5"
/protein_id="AAF71795.1"
/db_xref="GI:8052531"
/translation="MAVLSKPVAIKSGFSIPIVDMSDPSKHALVACEDFGFFKV
INHGVAELVSEIHEVDFSLPKSEKTQVAGYFGVNSKIGNGVDGVWVEYLLMN
ANHDGSGPLFPSSLKSGTFRNALLEYTSVRKMTFDVLEKTDGLGIXPRNLSKL
VSDQNTSILRLNHYVPCPLSNKNTGKNGVGFGEHTDPOIISVLRGNTSGQLNL
NDGSLVPPDHTSFEFVNGSLQVWNTNGRFSVHRVLANCKSRVSMVYFAGPSLT
QRIAPLTLCDNEDELVEFTWSEYKNSYNSRLSDNRLOQFERKTIKNLLN"
complement(join(15755..16715,16810..16829))
/notes="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F3f9.6"

```

```
/protein_id="AAF71812.1"
/db_xref="GI:8052548"
/translation="MPKRLGIRSELSPORSPRLSTLSLSDPHLGRPIITDRSP
KGLDRRSPRSGGHTDPSOKLGSRTSGLSOLGOAQBELRLKQOLAKAAAKKR
AOELRRPKSKKPPAPLSDQIPDCHQOETDVEVLEDEKAKSEKTNDELASKEDQ
INVKLARLYDLEKREVSLESENEELKQOLKKTDTMESCACAKADEIAKUSQVGBELE
ESNETAKLKLSVEBAKTLAEAKKLLKLVQTEQMKADAAAALVSGVENGRF
SQCSQMEKHFAGRVSPGWADSDGSGRKKMFKWKKKGQK"
complement(join(19162..19689,19767..19833,20044..20164,
20234..20378,20482..20623,20693..20738,20826..20877,
20965..21156))
/feature="unknown protein; similar to ESTs gb|t41672.1,
gb|AI992710.1, and gb|AA586078.1"
/codon_start=1
/evidence=not_experimental
/product="F3F9.7"
/protein_id="AAF71811.1"
/db_xref="GI:8052547"
/translation="MGNLGRKROVBERYTKPOGLYVNDVVKLRKLIVESKLAP
CYPDDESDHLEECPIFLYPSINRSCWKSICTEFLQMNPNASAPPTOKFLIN
LMGPPFVSSITFPRCPCKTPNVAVERGVSKSEKIEQVEQRIEAKINRQKE
MODDERKQKLESCESSSMTGEMEXGSGFLFVFWFAAISYNSIMDDGETAPQK
NASVYQSRGRNREDDVDLELMYMEALWLSVQBTGTORNSASGEITSSRYVT
DNHSYSPRPVPTVEPATFSSGSLSCAISALAEQMYGESSHNNHNVNVSYY
SNLPGNCDSDYDIEQVGDIDNHHRHRYHMETGSGNSYVSSYMTGEGFHPPPPP
PLVIVPESFEBQMMAMAVSMAEVHATTCAPTETWQ"
complement(join(22808..23063,23118..23244,23511..23706))
/feature="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F3F9.8"
/protein_id="AAF71810.1"
/db_xref="GI:8052546"
/translation="MSGRGKVKSEPMKVVFINTOVETDARSFKTVQELTGKNAIVA
AGPPSPAPDRCYDGSCKIPITNVFWKIMSLVSRPMSNPFKRAYSPFMLEK
RTMKPTNYQNKCSVCNKAUKATSRLSQASTMGFLSWGFAQVQLGITTFIFFI
YLYFMPFSCIMFLTKILSLKLVLSKV"
join(26752..26907,27020..27871,27923..28153)
/feature="similar to exopolylacturonase precursor
sp|Q00293|PGLX_ASPTU"
/codon_start=1
/evidence=not_experimental
/product="F3F9.9"
/protein_id="AAF71796.1"
/db_xref="GI:8052532"
/translation="WVSTISGFLILFVIAVASSISAAPSALVGRKVFVRSYGARG
DKTNTWATKAWDAQWGLPRVYFPGTFYLGVAFTGPCCKSRISFIKGTLLA
PKDPAIKODSWIIFRYVDLYTVSGGILDGSGYSWPLNCRQTHNCRALPMWGLQ
FVRFELRIKINSKMGHLNPFVQHEIDITVNIKAPGDSPTDGIKIGSNMKIH
HVDIATGDCIALLSGMNDINKVCGFPHGIGSVGLKFKGKSVQGLIVRSIFN
GTSNGVRIKTPSPPEPLVSNFLPNLQMDIVQSPINIDQRYCNPSPCSFQVSLTR
NKSFSKIDRYKFNINIGTSTAKEAVLQCKSNVPCXVQLFNINIVHRRDGPATS
VCENVGMGTGKISPPSCIR"
33566..33539
/feature="similar to 9-cis-epoxycarotenoid dioxygenase
gb|AAF26356.1"
/codon_start=1
/evidence=not_experimental
/product="F3F9.10"
/protein_id="AAF71797.1"
/db_xref="GI:8052533"
/translation="WTIIITISGMVIYLLSODAHHSOYQNTNLVLKLPKIPKOTAA
FNQESPMASTLLPSTQFLDPTSTSSSRPKLQSLSPSSTLRNKLVVPCVVS
SVNKSXSVSSLSQSTPFKPSWKLCNDVNLIPKTNQKPLNVQRTAAWLDAVE
NAMIHSHRRHPKPTADPAVQAGNFPPVPVPHNLPTVGTVECIQGVYVANGA
NPLHFKFVSGHLLFDGDMVHVRVNGSVSYACRFTNRLVQRECECRPVFFKAIGE
LHGLGIAKLMLFNTGLFGLVDPTGGLGVANAGLVYFNGHLLAMSEDDLPHYVKVTO
```

Query Match 47.9%; Score 862.8; DB 8; Length 95769;  
Best Local Similarity 73.1%; Pred. No. 6.2e-248;  
Matches 1139; Conservative 0; Mismatches 407; Indels 12; Gaps 2;  
245 ACATAACACAGATGAATTTGTTCCAGAGACCGCGCGGACGGTGGACGCGGGAGG 304

Db 35067 ACAGCAAAACACATTCAGAGTGTTTTCGTGAGATAGGCTAAACCTTAAACACAGGGAAT 35126  
 QY 1382 CAACCTGCGGTCGATCATCTCCAAACGAGATCAACAGCTACACTCGAAGCAGGATGG 1441  
 Db 35127 CAACACGTAGACGGGTATCTC-----GGAACAAGTTAATCTCGAAGCTGGTATGG 35177  
 QY 1442 TCAACGAACATATGCTCGGCGGTAAACCAAAATTCGCTTACTTTAGCGGACCGT 1501  
 Db 35178 TCAACCGGAATTTATTAGGTAGAAAAACCGGTATGCTTATCTGGCTTTAAACCGAACCGT 35237  
 QY 1502 GGCTAAAGTCTCAGGATTCGCTAAAGTTGATCTCACTACTGAGAGAGTTAAGAACATC 1561  
 Db 35238 GGCTAAAGTCTCAGGATTCGCTAAAGTTGATCTCACTACTGAGAGAGTTAAGAACATC 1561  
 QY 1562 TTTACGCGGATACCGGTTACGAGGAGAGCTCTGTTTCTCCCGGAGAGGAGGAGG 1621  
 Db 35298 TTTACGCGGAGGGAATACGAGGAGAGCTCTGTTTCTACCTTCGCTGACGAGAGAG 35357  
 QY 1622 AAGACGAGATACATCTCTGTTTCTGTTTCAACGAGAGAGATCGAATCGGAGTTAC 1681  
 Db 35358 AAGACGAGGTTACATTAATGTTGTTTCAACGAGGAGAGAGGTTGAATCGGAATCC 35417  
 QY 1682 AGATAGTTAAGCGGCTTAGCTTAGAGGTTGAAGCAACGGTTAAACTTCCGTCGAAGGTTTC 1741  
 Db 35418 AACTATAAGCGGTTAATAGAGCTTGAAGCTACCGTTACCGTTCCGTCGAGAGTGC 35477  
 QY 1742 CGTACGAGTTACGAGTACATTCATCGAGCGGATGATTTGGGAGACGAGTCTGTCG 1799  
 Db 35478 CTTACGCTTTCCAGGAAACGCTTATCAGTAAGGAAGATTTATCGAAGCAAGCTTTGTG 35535

RESULT 8  
 AX148316  
 LOCUS 1818 bp DNA linear PAT 08-JUN-2001  
 DEFINITION Sequence 15 from Patent EP1161794.  
 AX148316  
 VERSION AX148316.1 GI:14347203  
 KEYWORDS  
 SOURCE Lycopersicon esculentum (tomato)  
 ORGANISM Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; lamids; Solanales; Solanales; Solanaceae; Solanum; Lycopersicon.  
 1  
 Iuchi, S., Kobayashi, M. and Shinozaki, K.  
 Transgenic plants carrying neoxanthin cleavage enzyme gene  
 Patent: EP 1161794-A 15 18-JUL-2001;  
 Riken (JP)  
 Location/Qualifiers  
 1..1818  
 /organism="Lycopersicon esculentum"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:4081"  
 1..1818  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAC41199.1"  
 /db\_xref="GI:14347204"  
 /translation="MATTSHATNTWIKLSPSPKEFGFASNSISLLKNQNNROSL  
 NINSLQAPLHPKQSSNYQTPKNTTISHPKQNNNNSSSTSKWNLVKAAAL  
 DAVESALFKHELPKTPADPRVQISGNFAPVNPVQSLPVTGKIPKQVQVVR  
 NGAPLPSPTAGHFFFDGDMVHAFQKNGSAYACFTRELRLVQEKALRDPVPRKA  
 IGLHSGIARLMLFYARGLFLGVDHSGKGTGVANAGLVNENNELAMSDDLPYHVK  
 VTGDLKTEGRFDGQKSTMTAHKLPDVSSELFSALSDVYLQPKLYKFRSKNG  
 EKSNDVEIPEDPTNMHDFAITENPVIPOQVFKMSEMRGSPVYDKNKVRSG  
 ILDKYQDGLSKLWEPDPCFHLNWAEEADEIVIGSCMTPTDPSIFNECDGL  
 KSVLSEILNLTGSKTRKSIENPDQVNLNAGVNRNKLGRKTEYLAIAEPWPK  
 VSGPAKVLFTCEVEKFTYGDNKYGGELFLPRDPNSKEEDDGVILAFVHDEKWKSE  
 LQIVNMSIKLEATVKLPSRPVPGPHGTFINANDLANQA"  
 569 a 383 c 369 g 497 t

FEATURES  
 source

CDS

BASE COUNT  
 ORIGIN

Query Match

45.8%; Score 824.6; DB 6; Length 1818;

Best Local Similarity 69.6%; Pred. No. 8.4e-237;  
 Matches 1134; Conservative 0; Mismatches 489; Indels 6; Gaps 1;  
 QY 170 CTCAGAGCTCTTCATTTTCCCTTAAGCAATCATCAAACTCTCCGCGCATTTGTTGTTAAGCCCA 229  
 Db 185 CTTCAATTAATCAACACCAAGATTAATACATTTTACACCCCAAAACAGAAACAAACA 244  
 QY 230 AAGCAAAGAAATCAACACATCAACAGATGAATTTGTTTCCAGAGAGCGGCGGCGACGCT 289  
 Db 245 ACTCCTCTCTTCTTCAACTTCCAAGTGAATTTAGTGCAGAAAGCAGCAGCAATGCTCTT 304  
 QY 290 TGGACGGCGGCGGAGGTTTCTTGTTCAGCCACAGAGAGCTTACACCCGCTTCTTAAACCGG 349  
 Db 305 TAGATGCTGTAGAAGTGTCTTTAACTAAACATGAATTTGAAACACCTTTCCGGAACAG 364  
 QY 350 CTGATCTCTAGTGTTCAGATCGCCGGAATTTTGTCTCCGTTGATGAATGAACAGCCGTCGGC 409  
 Db 365 CCGACCCACGAGTCCAGATTTCTGGAAATTTTGTCTCCGTTACCGGAAATTTCCAGTCTGTCT 424  
 QY 410 GTATCTTCCGCTGGTTCGGAACCTTCCCGATTCATCAAGAGAGTGTATGTGCGCAACG 469  
 Db 425 AATCTTCTCCGTCACCGGAATAATACCAATGTGTTCAGGCTTTTACGTTCGAAGC 484  
 QY 470 GAGCTAACCCACTTCAAGAGCGGTTGACAGTCAACCTCTTTCGACGAGAGAGCGGTATGG 529  
 Db 485 GAGCTAACCCCTCTTTTGAACCAACCGCGGACACCACTTTCTTCGACGCGGACGCGTATGG 544  
 QY 530 TTCACGCGCTCAAAATTCGACACAGCTTTCAGTACTACGCTTTCGCGGTTTACTCAGACTA 589  
 Db 545 TTCACGCGCTTCAATTTCAAAATGGTTCGCTAGTGTACGCTTTCGCGTTTCACTGAACAG 604  
 QY 590 ACCGCTTTGTTTCAAGAACTCAATTTGGGTTCGACCGGTTTTCCTCCAAAGCCATCCGTTGAGC 649  
 Db 605 AGAGCTTTGTTTCAAGAAAGCTTTGGGTTCGCTTTCCTTAAAGCCATTTGTTGAAT 664  
 QY 650 TTCACGCGCACACCGGTTTTCGCGGCTCATCTTTCGACGCGAGCTTCGACCGGTA 709  
 Db 665 TACATGGTCACTCTGGAATTCGAAAGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCT 724  
 QY 710 TAGTCGACCCCGGCAACACGAAACCGGTGTAGCTAACCGCGGTTTGGTCTTTCATTCGCTTTCGCT 769  
 Db 725 TTGTTGATCAGACTAAAGAACTGGTGTTCGAAACCGCGGTTTTCGCTTTCATTCATTAACC 784  
 QY 770 GGTATTGGCTATTCGCGAGGATGATTTACCTTACCAAGTTTCAGATCACTCCCAATGAG 829  
 Db 785 GATTACTTGTATGTCTGAAGATGATTTGCTTTCATGTAAGGTAACACCCACCGCG 844  
 QY 830 ATTTAAAAACCGTTGCTGCTTCGATTTTTCGATTCGCAATTTAGAAATTCACCAATGATTCGCC 889  
 Db 845 ATCTTAAACACAGAGGTTCGATTCGATTCGACGCGGCTAAATTCACCAATGATGATGCTC 904  
 QY 890 ACCGAAAGTCGACCGGAAATTCGCTGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCT 949  
 Db 905 ACCCAAGCTCGACCCGAGTTTCGCGTGAGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCT 964  
 QY 950 AGCTTTTACCTAAATTAATTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCT 1009  
 Db 965 AGCCATACCTCAAGTACITTCAGATTTTCAAAAAATTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCT 1024  
 QY 1010 TTCAGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCT 1069  
 Db 1025 TTCCAGTTGAAGACCCCAACATGATGATGATTTTCGCAATTTACTGAGAACTTCGCTGCTCA 1084  
 QY 1070 TACCTCAGCAGCAAGTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCT 1129  
 Db 1085 TTCTCTGATCAACAGTCGTTTTCAGATGTCGTAATGATCGTGGAGGTTTCAACCGGTTG 1144  
 QY 1130 TTTTACGACAGAAACAGGTCGCAAGATTCGCGGATTTTACAGAAATTCGCGGAGGATTCAT 1189  
 Db 1145 TTTTACGACAGAAACAGGTTTTCGCGATTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCT 1204  
 QY 1190 CGAACATTAAGTGGATGATGCTCCAGATTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCT 1249

Db 1205 CTGATTGAAATGGGTGAAGTACCTGATGTTCTGTTTCCACCTCTGGAATGCTTGGG 1264  
Qy 1250 AAGAGCCAGAAACAGATGAAGTCGTGATAGGTCCTGTATGACTCCACAGACTCAA 1309  
Db 1265 AAGAGCAGAAACAGATGAAGTCGTGATAGGTCCTGTATGACTCCACAGACTCAA 1324  
Qy 1310 TTTTCAAAGAGTCTCAGAGAAATCTCAGAGAGTCTCTGTGAAATCCGCTGAATCTCA 1369  
Db 1325 TTTTCAATGAATGTGATGAAGGGCTAAAGAGTGTATTCGGAATCCGCTCAATTTGA 1384  
Qy 1370 AAACCGGTGAATCAACTCCGCTCCGATCATCTCCAAAGAGATCAACAAGTCAACCTCG 1429  
Db 1385 AAACAGGGAATCAACAGAAATCCATATCGAAACCCGATGAAACAAGTGAATTAG 1444  
Qy 1430 AAGAGGGATGTCAACAGAAATCTGTCGGCCGCTGAAACCAAAATTCGCTTACTTGGCTT 1489  
Db 1445 AAGCTGGAATGTGAAACCAAACTCGGAAGGAAACAGATGATGCTTATTTGGCTA 1504  
Qy 1490 TAGCCGAGCGGTGGCTAAAGTCTCAGGATTCGCTTAAAGTTCATCTCACTACTTGAGAG 1549  
Db 1505 TCGCTGAACCATGGCCAAAGTTCGTTTGGTAAAGTAAACCTGTCTACCGTGAAG 1564  
Qy 1550 TTAAGAAACATCTTTACGCGGATAACCGTTACGAGGAGAGCTCTGTTTCTCCCGGAG 1609  
Db 1565 TTGAGAAATTCATTTATGTGCAACAATAATGTTGGGGGAACCTCTTTTTTACCAGAG 1624  
Qy 1610 -----AAGGAGGAGAGAGAGGATACATCTCTGTTTCTGTTTCTACGAGCAGAGA 1663  
Db 1625 ACCCAACAGCAAGAGAGAGAGTGGTTATTTTATTTAGCTTCTGTTTACCATGAGAAAG 1684  
Qy 1664 CATGGAATCGGAGTTACAGATAGTTAAGCCGCTTAGCTTAGAGTTGAAGCAACGTTA 1723  
Db 1685 AATGGAATCAGAACTGCAAAATGTTTAAAGCAATGAGTTTGAAGTTGAGGCAACTGTGA 1744  
Qy 1724 AATCTTCGTCAGGTTCCGTACGATTTACGATTTACGATTTACGAGCCGATTTGG 1783  
Db 1745 AGCTTCCATCAGAGTTCCTTATGATTTTATGGAACATTCATAAACCCCAATGATTTGG 1804  
Qy 1784 CGAAGCAGG 1792  
Db 1805 CAATCAGG 1813

BD017436 1818 bp DNA linear PAT 27-AUG-2002  
Transgenic plant using neozanthine cleaving enzyme gene.  
BD017436  
BD017436 1 GI:22558612  
JP 2001258579-A/8.  
Lycopersicon esculentum (tomato)  
Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.  
1 (bases 1 to 1818)  
Tsuchi, K., Kobayashi, M. and Shinozaki, K.  
Transgenic plant using neozanthine cleaving enzyme gene  
Patent: JP 2001258579-A 8 25-SEP-2001;  
THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH  
OS Lycopersicon esculentum (tomato)  
PN JP 2001258579-A/8  
PD 25-SEP-2001  
PF 11-JAN-2001 JP 2001003476  
PI KITOCHI TUCHI, MASATOMO KOBAYASHI, KAZUO SHINOZAKI PC  
C12N15/00, A01H5/00, C12N9/02, C12N15/00, C12N5/00 CC  
Transgenic plant using neozanthine cleaving enzyme gene PH Key  
Location/Qualifiers  
FT CDS Location/Qualifiers  
1..1818  
/organism="Lycopersicon esculentum"  
/mol\_type="genomic DNA"

BASE COUNT 569 a 383 c 369 g 497 t  
ORIGIN  
Query Match 45.8%; Score 824.6; DB 6; Length 1818;  
Best Local Similarity 69.6%; Pred. No. 8.4e-237;  
Matches 1134; Conservative 0; Mismatches 489; Indels 6; Gaps 1;  
Qy 170 CTCAGCTCTTCAATTTCCCTAAGCAATCATCAAACTCTCCGCCATTTGTTTAAAGCCCA 229  
Db 185 CTTCAAATATCAACACCAAGATATATCAATTTTCAACCCCAAAACAGAAACAACA 244  
Qy 230 AAGCCAAAGATCCACACATTAACAGATGAATTTGTTCCAGAGAGCGGCGGAGCGT 289  
Db 245 ACTCTCTCTTTCTTCAACTTCCAGTGGAAATTTAGTCAGAAAGCAGAGCAATGCGCTT 304  
Qy 290 TGGACGCGGCGGAGGTTTCTTTGTCCAGCAACAGAAAGTACACCCGCTTCTCTAAACCGG 349  
Db 305 TAGATGCTGTAGAAAGTGTCTTTAACTAAACATGAACCTTGAACACACCTTTGCCGGAACAG 364  
Qy 350 CTGATCTAGTGTTCAGATCGCGGAAATTTGCTCCGGTGAATGAACAGCCGCTCCGCG 409  
Db 365 CCAGCCACAGTCCAGATTTCTGGGAATTTGCTCCGGTACCGGAAATTCAGTCTGTC 424  
Qy 410 GTAACTCTCCGCTGCGGAAACCTTCCGATTTCCATCAAGAGAGTGTATGTGCGCAACG 469  
Db 425 AATCTCTCCGCTACCGGAAATATACCAAAATGTGTTCAAGGCGTTTACGTTTCAAAACG 484  
Qy 470 GAGCTAACCACTTCAGAGCCGTCAGAGTCAACACTTCTTCGAGAGAGCGGTATGG 529  
Db 485 GAGCTAACCCCTCTTTTGAACCAACCGCGGACACCATTTCTTCGACGCGGAGCGGTATGG 544  
Qy 530 TTCACGCGCTCAAAATTCGAACACCGTTTCAGCTAGCTTACGCTTCCGCGTTTACTCAGACTA 589  
Db 545 TTCAGCGCTTCAATTCMAAAATGGTGGCTAGTTAGCTTGGCTTTCACTGAAACAG 604  
Qy 590 ACCGTTTGTTCAGAGACGTCATTTGGGTGCGACCGGTTTTTCCCAAGACATCGGTGAGC 649  
Db 605 AGAGGCTTGTTCAGAAAGAGCTTTGGGTGCGCCTTTTCCCTAAAGCACTTGGTGAAT 664  
Qy 650 TTCACGCGCACACCGGTATTGCGGACTCATGTATTTCTACGCGCAGAGCTCGAGCCGGTA 709  
Db 665 TACATGCTCACTTGGAAATGCAAGGCTTATGCTGTTTACGCTCGTGGGCTCTTCGGAC 724  
Qy 710 TAGTCGACCGCGCACCGAAACCGGTGTAGCTTAACCGCGTTTGGTCTATTTCATATGGCC 769  
Db 725 TTGTTGATCAGCTAAAGGAACTGGTGTTCGAAACCGCGGTTTAGTCTATTTCATTAACC 784  
Qy 770 GGTATTGGCTATGTGCGAGATGATTTACCTTACCAAGTTCAGATCACTCCCAATGGAG 829  
Db 785 GATTACTTGTCTATGCTGAAGATGATTTGCCCTTACCATGTAAGGTAACACCCCGCGG 844  
Qy 830 ATTTAAAAACCGTTGGTTCGGTTCGATTTTGTATGGACAATTTAGAAATCCACAATGATGCC 889  
Db 845 ATCTTAAACAGAGGGTTCGATTCGATTTCCAGCGCCAGCTTAAATCCACATGATAGCTC 904  
Qy 890 ACCGGAAGTCGACCGGAAATCCGGTGAACCTTCTGCTTTAGCTACGAGCTGTTTCAA 949  
Db 905 ACCCAAGCTCGACCGGAGTTTCCGGTGAGCTATTGCTCTTTAGCTACGATGATTCAGA 964  
Qy 950 AGCTTACCTTAAATATCTTCCGATTTCTCAGCGGACGGAACCTAAATCACCAGAGCTCGAGA 1009  
Db 965 AGCATACCTCAAGTACTTCCAGATTTTCAAAAAATGGGAAAAATCAAATGATGTTGAA 1024  
Qy 1010 TTCAGCTTGTATCGCAACAGATGATGACGATTTTCGGATTACAGAGAACTTCTGCTCG 1069  
Db 1025 TCCAGTTGAAGACCCCAACCAATGATGATGATTTTCGCAATTTACTGAGAACTTCTGCTCA 1084  
Qy 1070 TACTGTACCAAGAGTGTGTTTCAAGCTGCGGAGATGATCCCGGTTGGGTCTCCGCTGG 1129  
Db 1085 TTCTGTATCAACAGTCTGTTTCAAGATGCTGAATGATCCGTGGAGGTTTCAACCGGTGG 1144  
Qy 1130 TTTACGACAAGAACAGGTCGCAAGATTCGGGATTTTAGACAAATACGCGGAGATTCAT 1189

REMARK  
REFERENCE 3 (bases 1 to 2171)  
Direct Submission  
TITLE Burbridge A., The University of Nottingham,  
Physiology and Environmental Science, Sutton Bonington Campus,  
Loughborough, Leicestershire, LE12 5RD, UK  
COMMENT On Jan 13, 1998 this sequence version replaced gi:2243153.  
FEATURES  
Location/Qualifiers  
1..2171  
/organism="Lycopersicon esculentum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4081"  
79..1896  
/codon\_start=1  
/product="nine-cis-epoxycarotenoid dioxygenase"  
/protein\_id="CA810168.1"  
/db\_xref="GI:2769642"  
/translation="MATTTSHTATNWKTKLMPSSKEFGFASNSILKXNHRNSL  
NINSSLOAPILHPKQSSNYQTPKNTTSHPKQENNSSSSTKWNLVQKAAVAL  
DAVESALPTKELEHPLPKTADPRVQISGNFAPVPENPCQSLPTQKIPKQGVYVR  
NGANPLPEPTAGHHFFPDGDMHVAQVQNGSASVACRFTETELVKALGRVFFKVA  
IGELHSGGTAEMLFVARGLFGLVHDSKGTGVANAGLVYFNRLKMSDDLPYHVK  
VTFGDLKTEGDFDQGLAKSTMAHPKLDPVSGELFALSVDVIQKPYLYKPKVSKNG  
EKNDVBEIPVEDPTMHDPFALTENFVVIPOQVYFPMSEMRGSGSVVVDKXVSRG  
ILDKYAKGSLKXWVEPDCFCFHLNMAEAEETDELIVIGSCMPEDSIFNCEDEGL  
KSLYSEIRLNLTKGSTRKSIIEPDEQVNLKAGMVRNKLGRKRTAYLALAEWPK  
VSGFANVLTGEVKEIYGDNKYGGEPFLPRDPNSKEEDDGYILAFVHDEKEWKE  
LQIVNANSLKLEATVKLPSRPVYGFHGTFINANDLANQA"  
BASE COUNT 688 a 435 c 431 g 617 t  
ORIGIN  
Query Match 45.8%; Score 824.6; DB 8; Length 2171;  
Best Local Similarity 69.4%; Pred. No. 8.7e-237;  
Matches 1134; Conservative 0; Mismatches 489; Indels 6; Gaps 1;  
QY 170 CTCAGCTCTTCATTCCTTAAGCAATCATCAAACTCTCCGCCATTTGTTTAAAGCCCA 229  
Db 263 CTTCAATATCAACACCAAGAAATATATCAATTTACACCCAAACAAACAAACA 322  
QY 230 AAGCCAAAGAAATCCAAACTAAACAGATGAATTTGTTCCAGAGAGCGCGCGCAGGT 289  
Db 323 ACTCCTCTCTCTTCTTCAACTTCCAACTTGAATTTAGTGCAGAAAGCAGCAATGGCTT 382  
QY 290 TGGACGCGCGGAGGGTTTCTTGTACGACGAGAGTACACCGCTTCTTAAACCG 349  
Db 383 TAGATGCTGTAGAAAGTCTTTAACTAAACATGAATTTGTTGCGGAAACCGTTCGGAACAG 442  
QY 350 CTGATCTCTAGTGTTCAGATCGCGGAAATTTTGTCCGGTGAATGAACACAGCCCGTCCGCG 409  
Db 443 CCGACCCACGAGTCCAGATTTCTGGGAATTTTGTCCGGTACCGGAAATCCAGTCTCTC 502  
QY 410 GTAATCTTCGGTGTTCGAAACTTCCGATTCATCAAGAGGTATGTGGCAACG 469  
Db 503 AATCTCTTCGGGTTCACCGGAAATTTACCAATGTGTTCAAGGCGGTTCAGTTCGAAACG 562  
QY 470 GAGCTAAACCCACTTCCAGAGCGGTGACAGGTGACCACTTCTTCGACGAGACGTATGG 529  
Db 563 GAGCTAACCTCTTTTGAACCAACCGCGGACACCAATTTCTTCGACGCGACGTATGG 622  
QY 530 TTACGCGCGTCAAAATCGAACAGGTTTACGCTAGTACGCTTCCCGGTTTACTCAGACTA 589  
Db 623 TTACGCGCGTCAAAATCGAAATTTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 682  
QY 590 ACCGGTTTCTTCAGGAACCGTCAATTTGGGTGCGACCGGTTTCCCAAAAGCCATCGGTGAGC 649  
Db 683 AGAGGCTTGTTCAGAAAGGCTTTGGGTGCGCTGTTCCTTAAAGCCTTGTGTGAAT 742  
QY 650 TTACGCGCGCACCGGTATTCGCGACTATTCGCTATTCACGCCAGAGTCGACCGGTA 709  
Db 743 TACATGCTCAGCTGGAATTCGAGGCTTATCTGTCTTTTACGCTCGTGGGTCTTTCCGAC 802

1145 TTTTACGACAGAACAAAGTTTCCGATTTGGTATTCTGGATAGTACGCGAAAGATGGGT 1204  
QY 1190 CGAACATTAAGTCGATGATGTCACAGATTGCTTCTGCTTCCATCTCTGGAACGCTTGGG 1249  
Db 1205 CTCGATTTGAATGGGTGAGTACCTGATGTTTCTGTTTCCACCTCTGGAATGCTTGGG 1264  
QY 1250 AAGAGCCAGAAACAGATGAAGTCGTCGATGATGGTCTCTGATGACTCCACAGACTCAA 1309  
Db 1265 AAGAGCAGAAACAGATGAATCGTTGTAATTGGTTTATGATGATGATGATGATGATGATG 1324  
QY 1310 TTTTACGACGATCTGACGAGATCTCAAGAGTGTCTGCTGAAATCCGCTGAAATCTCA 1369  
Db 1325 TTTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1384  
QY 1370 AAACCGGTGAATCAACTCGCGCTCGATCATCTCCAAAGAAATCAACAAAGTCAACCTCG 1429  
Db 1385 AAACAGGGAATCAACAAAGAAATCCATATCAATCAAGAAACCGGATGAACAAAGTGAAT 1444  
QY 1430 AAGCAGGATGGTCAACAGAAACATGCTCGCGCTGAAACCAATTCGCTTACTTGGCTT 1489  
Db 1445 AAGCTGAATGGTGAACCGAAACAACTCGGAAGGAACAGAGTATGCTTATTTGGCTA 1504  
QY 1490 TAGCCGAGCGTGGCTTAAAGTCTCAGGATTCGCTTAAAGTGTATCTCACTACGAGAG 1549  
Db 1505 TCGCTGAACCATGCGCAAAAGTTCTCGTTTTCGAAAGTAAACCTGTTTCAACGCTGAAG 1564  
QY 1550 TTAAGAAACATCTTACCGGATTAACCGTTACGAGAGAGAGCTCTGTTTCTCCCGGAG 1609  
Db 1565 TTGAGAAATTCATTTAAGTGACAAACAAATATGTTGGGAACTCTTTTTCACCAAGAG 1624  
QY 1610 -----AAGGAGGAGAGAGAGAGAGATACATCTCTCTGTTTCTGTTACGAGAGAGAG 1663  
Db 1625 ACCCCACAGACAGAGAGAGAGAGATGTTTATTTTATTTTATTTTATTTTATTTTATTT 1684  
QY 1664 CATGGAATCGGATTCAGATAGTTTAAACCGCTTACCTTAGGTTGAGGCAACGCTTA 1723  
Db 1685 AATGGAATCAGAACTGCAATTTTAAACGAAATGATTTGAAATTTGAGGCAACGCTTA 1744  
QY 1724 AACTTCGCTCAAGGTTCCGTTACGAGATTCACGTTACATTCATCGGAGCGAGATTTGG 1783  
Db 1745 AGCTTCCATCAGAGTTCCITATGATTTATGAAATTCATGAAATTCATGAAATTCATG 1804  
QY 1784 CGAAGCAGG 1792  
Db 1805 CAAATCAGG 1813  
RESULT 10  
LENEOXANT 2171 bp mRNA linear PLN 22-MAY-1998  
DEFINITION Lycopersicon esculentum mRNA for nine-cis-epoxycarotenoid  
dioxygenase.  
ACCESSION Z97215  
VERSION Z97215.1 GI:2769641  
KEYWORDS nine-cis-epoxycarotenoid dioxygenase.  
SOURCE Lycopersicon esculentum (tomato)  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.  
1 Burbridge, A., Grieve, T.M., Jackson, A., Thompson, A. and Taylor, I.B.  
2 Structure and expression of a cDNA encoding a putative neoxanthin  
cleavage enzyme (NCE) isolated from a wilt-related tomato  
(Lycopersicon esculentum Mill.) library  
J. Exp. Bot. 47, 2111-2112 (1997)  
REFERENCE  
Burbridge, A.  
Direct Submission  
TITLE Burbridge A., The University of Nottingham,  
Physiology and Environmental Science, Sutton Bonington Campus,  
Loughborough, Leicestershire, LE12 5RD, UK





/evidence=experimental		
repeat_region	8285..8575	
/rpt_family="tomato anionic peroxidase inverted repeat (TAPiR)"		
repeat_region	10047..10285	
/rpt_type=INVERTED		
/notes="putative"		
repeat_region	10047..10063	
/transposon="stowaway element"		
repeat_region	10047..10063	
/rpt_type=INVERTED		
/rpt_type=TERMINAL		
repeat_region	10359..10756	
/notes="putative mobile element"		
BASE COUNT	6823 a 2631 c 2645 g 6919 t	
ORIGIN		
Query Match		
Best Local Similarity 45.8%; Score 824.6; DB 8; Length 19018;		
Matches 1134; Conservative 0; Mismatches 489; Indels 6; Gaps 1;		
Qy	170 CTCAGCTCTTATCCCTTAAGCAATCATCAAACTCTCCGCCATTTGTTAAGCCCA	229
Db	5666 CTTCAAAATTATCAACACCAACAGATAATAACAATTTACACCCAAAACAAGAAACAACA	5725
Qy	230 AAGCAAGAAATCCACACTAAACAGATGAATTTGTCAGAGAGCGCGGCGAGCGT	289
Db	5726 ACTCCTCTCTCTCAACTTCCAACTTCAAGTGAATTTAGTCAGAAAGCAGCAATGGCTT	5785
Qy	290 TGAACGCGCGAGGGTTTCCTTGTGAGCCACAGAAAGCTACACCGCTTCTTAACCG	349
Db	5786 TAGATGCTGTAGAAAGTGTCTTAACATAACATGAACCTTGAACACCCCTTTGCCGAAACAG	5845
Qy	350 CTGATCCTAGTGTTCAGATCCGCGAAATTTCTCGGTGTAATGAACACCGCTCCGCG	409
Db	5846 CCGACCCACAGTCCAGATTTCTGGGAATTTTCTCGGTACCGGAATTCAGTCTGTCT	5905
Qy	410 GTAATCTCCGGTGTTCGAAAACTTCCCGATTCATCAAGAGTGTATGTGCGCAAG	469
Db	5906 AATCTCTCCGGTCCCGGAAAAATACCAATGTGTTCAGCGCTTTACGTTCCGAAACG	5965
Qy	470 GAGTAACCCACITCAGACCGGTGACAGTCAACCTTCTTCGCGAGAGCGGTATGG	529
Db	5966 GAGTAACCCCTCTTTTGAACCAACCGCCGGAACCACTTTCTTCGACGCGACGATGG	6025
Qy	530 TTCACGCGCTCAATTCGAAACACCGTTTCAGTGTACGCTTCCCGGTTTACTCAGACTA	589
Db	6026 TTCACGCGCTCAATTCGAAATTCGCTGCTGTACGCTTCCGCTTCTCACTGAAACAG	6085
Qy	590 ACCGGTTGTTCAGAAAGCTCAATGGTTCGACCGGTTTCCCAAGCCATCGGTGAGC	649
Db	6086 AGAGGCTGTTCAGAAAGAAAGCTTTGGTTCGCGCTGTTTCCCTTAAAGCCATTTGGTGAAT	6145
Qy	650 TTCACGCGCACACCGGTATTCGCGACTCATGCTTATTCAGCCAGAGCTGCAGCCGGTA	709
Db	6146 TACATGCTCATCTGGAATTCAGAGCTTATGCTGTTTACGCTCGTGGGCTCTTCGGAC	6205
Qy	710 TAGTCGACCCGGCACACGGAACCGGTGTAGTCAACCGGGTTTGGTCTATTTCAATGGCC	769
Db	6206 TTGTTGATCACAGTAAGAAAGAACTGGTGTGAAACCGCGGTTTGTCTATTTCAATAACC	6265
Qy	770 GGTATTGCTGTATGCGAGAGATGATTTACCTTACCAAGTTTCAGATCACTCCCAATGGAG	829
Db	6266 GATTACTTGTCTGTGAAGATGATTTGCTTACCATGTAAAGTAAACACCCACCGCG	6325
Qy	830 ATTTAAAAACCGTGGTGGTTCGATTTTGTATGAGCAATTAGAATCCCAATGATTCGCC	889
Db	6326 ATCTTAAAAACAGAGGTGATTCGATTTTCGACGCGCAGCTAAAAATCCACCATGATGTC	6385
Qy	890 ACCCGAAGTGCACCCGGAATCCCGTGAACCTTTCGCTTTAAGCTACAGCTCGTTTCAA	949
Db	6386 ACCCAAGCTGCACCCAGTTTCCGGTGAAGCTATTTGCTTCTAGTACGATGATTCAGAA	6445
Qy	950 AGCCTTACCTAAAATACTTCGGAATTCACCGGACGGAACTAAATACCGGACGCTCGAGA	1009

Db	6446 AGCATACCTCAAGTACTTTCAGATTTTCAAAAAATCGGAAAAATCAATGATGTTGAAA	6505
Qy	1010 TTCACTGTATGATCAGCAACAGCATGATGATGATTCGCGATTACAGAGAACTTCGTCGTCG	1069
Db	6506 TTCCAGTTGAAGACCCCAACAATGATGATTCGCAATTAATGAGAACTTCGTCGTCG	6565
Qy	1070 TACTGACCAAGCAAGTCTTTTCAAGCTCCGAGATGATCCGCGTGGGTCTCCGCTGG	1129
Db	6566 TTCTGTATCAACAAGTCTTTTCAAGATGATCCGATGATCCGCTGGGTTCACCGCTGG	6625
Qy	1130 TTTACGACAAGAAACAAGTCTCGAAGATTTCCGGATTTTAGACAAATACGCCAAGATTCAT	1189
Db	6626 TTTACGACAAGAAACAAGTCTTCGGAATTTCCGATTTTGGATAAGTACGCCAAGATCGGT	6685
Qy	1190 CGAATTAAGTGGATGATGATTCAGATTCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1249
Db	6686 CTGATTTTGAATGGTGTGAAGTACCTGATTTGTTTCTGTTTCTTCCACCTCTTGAATGCTTGG	6745
Qy	1250 AAGAGCCAGAAAACAGATGAAGTCTCTGATAGGCTCTCTGATGACTTCCACCAAGACTCAA	1309
Db	6746 AAGAGCAGAAAACAGATGAAGTCTCTGATGACTTCCACCAAGACTCAA	6805
Qy	1310 TTTTCAACGAGTCTGACGAGAACTCAAGATGCTCTGATGACTTCCACCAAGACTCAA	1369
Db	6806 TTTTCAATGAATGATGAAGGCTAAAGAGTGTTTTATCCGAAATCCGTTCTCAATTTGA	6865
Qy	1370 AAACCGGTGAATCAACTCGCGCTCCGATCATCTCCAAACGAAGATCAACAAGTCAACCTCG	1429
Db	6866 AAACCGGGAATCAACAAGAAATCCATATCGAAACCCCGATGAACAAGTGAATTTAG	6925
Qy	1430 AAGCAGGATGGTCAAACAGAAACATGCTCGCGCTAAACCAAAATTCGCTTACTTGGCTT	1489
Db	6926 AAGCTGGAATGGTGAACCGGAAACAACTCGGAAAGAAACAGAGTATGCTTTATTTGGCTA	6985
Qy	1490 TAGCGGACCGCTGCGCTTAAAGTCTCAGATTCGTTAAAGTTGATCTCACTACTCGGAAAG	1549
Db	6986 TCGTGAACCAAGGCGCAAGATTTCTGGTTTGAAGAAATTAACCTGTTCAACCGTGAAG	7045
Qy	1550 TTAAGAAACATCTTACCGGATTAACCGTTACGAGGAGAGCTCTCTCTCTCTCTCTCTCTCT	1609
Db	7046 TTGAGAAATCATTTATGTTGACAAACAATATGTTGGGGAACCTCTTTTATACCAAGAG	7105
Qy	1610 -----RAGGAGGAGAGAGAGAGAGAGATACCTCTGTTTCGTTTCAGCAGCAGAGA	1663
Db	7106 ACCCAACAGCAAGAGAGAGAGAGATGTTATATTTAGCTTTCGTTTCAGCAGAGAGAAG	7165
Qy	1664 CATGGAATCGAGTTACAGATAGTTAAACCGCTTAGCTTAGAGGTTGAAGCAACGGTTA	1723
Db	7166 AATGGAATCAGAACTGCAATTTGTTAACGCAATGAGTTTGAAGTTGGAGGCAACTGTGA	7225
Qy	1724 AACTTCCGTCAGGGTTCGTTACGATTTTCAAGTATTCAGGATTCATCGGAGCCGATGATTGG	1783
Db	7226 AGCTTCCATCAAGAGTTCCTTATGATTTTCATGAAACATTCATTAACGCCAATGATTGG	7285
Qy	1784 CGAAGCAGG 1792	
Db	7286 CAATCAGG 7294	

STU276244 2164 bp mRNA linear PIN 06-MAR-2000  
Solanum tuberosum mRNA for putative 9-cis-epoxycarotenoid  
dioxygenase (nced1 gene).  
STU276244 1 GI:7209268  
9-cis-epoxycarotenoid dioxygenase; nced1 gene.  
Solanum tuberosum (potato)  
Solanum tuberosum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; lamiales; Solanales; Solanaceae; Solanum.



REFERENCE 1  
AUTHORS Burbidge A., Taylor, I.B. and Thompson, A.  
TITLE Potato putative 9-cis-epoxycarotenoid dioxygenase 1 cDNA  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2164)  
AUTHORS Burbidge A.  
TITLE Direct Submission  
JOURNAL Submitted (02-MAR-2000) Burbidge A., Plant Science Division, The University of Nottingham, Sutton Bonington Campus, Loughborough, Leicestershire, LE12 5SD, UNITED KINGDOM  
FEATURES  
source  
1..2164  
/organism="Solanum tuberosum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4113"  
1..2164  
/gene="ncedi"  
73..1887  
/gene="ncedi"  
/function="cleavage of 9-cis-epoxycarotenoids"  
/codon\_start=1  
/product="putative 9-cis-epoxycarotenoid dioxygenase"  
/protein\_id="CAB76920.1"  
/db\_xref="GI:7209269"  
/translation="MATTTSHTATNTWIKPLSMESPSSKEFGFASNSISLLKNOHNRSLNINSLQAPLILHFPKQSNITQTPKTSIHFKQENNNSSISKNLVLKQAAAMLD  
AVEGALTKEHELEHLPKTDPRVQISGNFAPENPVQSLPVTGKIPKCVQGVYVRN  
GANPLFETAGRHFFDGGDMVHAFQFKNGSASACRPTETEFVQEKALGRPPFKAI  
GELHSGIATRLMFLPYARGLFGLIDHSRGTVANAGLVYFNRLLEMSDDLPHYKIV  
TPDGLKTEGDFDGLKMTIAHPKLDVQSGELFALSVDVIOKPYLKYFRSKNGE  
KSNDEIVEEPTWMDHFAITEKFVIFPDQVFPKSEMRIGSPVVDKNNYRFGI  
LDKYAKOSDLKWEVPCPCFHLNWEETEDIEIVIGSCMTPPDSIINECDEGLK  
SVLSIRLNLTKGSTRKAIENDEQVNLKGMVNRNKGRTQAYLALAEFPWKV  
SGFAVDLTFEVEKFTYGDNKGYPELFLPRDPSNKEEDDGYILAFVHDEKEWTSLS  
QIVNMTLKLEATVKLPRVPYFGHGTFINANDLANQA"  
BASE COUNT 667 a 437 c 433 g 627 t  
ORIGIN  
Query Match 45.0%; Score 809.4; DB 8; Length 2164;  
Best Local Similarity 70.7%; Pred. No. 3.4e-232;  
Matches 1093; Conservative 0; Mismatches 446; Indels 6; Gaps 1;  
QY 254 AGATGAATTTGTTCCAGAGAGCGGGCGGCGAGCGGTTGAGCGCGGAGGGTTTCCTTG 313  
DB 338 AGTGAATTTAGTGCAGAAAGCAGCAGCAATGGCTTTAGATGCTGTGAAGGTGCTTTAA 397  
QY 314 TCAGCCACGAGAGCTACACCGCTTCTTAAACCGCTGATCCTAGTGTTCAGATCGCG 373  
DB 398 CAAAACACGAACTCGAGCACCTTTGCCGAAAACAGCCGACCCAGGTCCAGATTTCTG 457  
QY 374 GAAATTTTCCTCGGTGTAATGAACAGCCGTCGGCGTAATCTTCGGGTGTCGGAAC 433  
DB 458 GGAATTTTCCTCGGTACCGGAAACCCAGTTTGCCATCTCTTCGGTACCGGAAA 517  
QY 434 TTCCGATTCATCAAGAGGTGTATGTGCGCAACGAGAGCTAAACCACTTCAGAGCGCG 493  
DB 518 TACCCAAATGTGTTCAAGCGGTTTACGTTTCAAGAACGAGCTAACCCCTCTTTTCAACCAA 577  
QY 494 TGACAGCTACCACTTCTTCAGCGGAGACGGTATGTTTCAGCGCTCAAAATTCGAACACG 553  
DB 578 CGCGCGGACCGCATTTCTTCAGCGGACGGATGTTGTTTCACGCCGTTCAATTCAAAATG 637  
QY 554 GTTCAGCTAGCTACGCTTCGCGGTTTACTCAGACTAACCGGTTTGTTCAGGAACGTCAAT 613  
DB 638 GGTCACTAGTTACGCTTCCCGGTTTCACTGAAACAGAGAGGTTTGTTCAGAAAAGGCTT 697  
QY 614 TGGGTGACCGGTTTTCCCAAGCCATCGGTGAGCTTCAGCGGCACACCGGTTATGCCC 673  
DB 698 TGGGTGCGCCGTGTTTCCCAAGCCATTTGGTGAATATCATGTGTCACCTCTCGAATTCGAA 757  
QY 674 GACTCATGTATTCTACGCCAGAGTGCAGCGCGGTATAGTCGACCCGCGCACACGAAACCG 733  
DB 758 GGTATATGCTGTTTACGCTCGTGGCTCTTCGGACTTATTGATCAGTAGAGGAACTG 817

QY 734 GTGTAGCTAACCCCGGTTTGGTCTATTCAATGGCCGGTTATTGGCTATTGTCCGAGGATG 793  
DB 818 GAGTTGCAAAACCGCGGTTTAGTCTATTTCATTAACCGATTACTTCTGTCTGAAGATG 877  
QY 794 ATTACCTTACCAAGTTTCAGATCATCCCAATGGAGATTTAAACCCGTTGTGTCGGTTCG 853  
DB 878 ATTTCGCTTACCATGTAAGGTAAACCCACCGCGGATCTTAAACACAGAGGTCGATTTCG 937  
QY 854 ATTTTGTATGCAATTTAGAAATCCCAATGATTGCCCCACCGAAAGTCGACCCCGGAATTCG 913  
DB 938 ATTTGACGGCCAGCTAAATCCCATGATAGTCTACCCAAAGCTCGACCCAGTTTCG 997  
QY 914 GTGAATCTTTCCTTTTAAGCTACGACGTCGTTTCAAGCTTAAAGCTTAACTTAACTTCCGAT 973  
DB 998 GTGAGCTATTTCCTTCTTAGCTACGATGTGATTTCAGAAGCCATACCTCAAGTATTTCAGAT 1057  
QY 974 TCTCCCGGACGGAATAAATCACCGGAGCTCGAGATTCAGTTGATCAGCTGATCAGCCAAAGTGA 1033  
DB 1058 TTTCAAAAATGCGGAAAATCAATGATGTTGAATCCAGTTGAGACCCCAACAATGA 1117  
QY 1034 TGCAGATTTTCGATTTACAGAAATTCGTCGTCTGCTACCTGACCGAAGTCGTTTTCAT 1093  
DB 1118 TGCATGATTTTCGCAATTTACTGAGAAATTCGTTCATCATCTCTGATCAACAAGTCGTTTTCAT 1177  
QY 1094 AGCTCCCGGAGATGATCCGCGTGGGTCTCCGCTGGTTCACGACGAACCAAGTCGCA 1153  
DB 1178 AGATGCTCGAAATGATCCGTCGAGGTTTACCGGTCGTTTACGACGAACCAAGTCGCC 1237  
QY 1154 GATTCGGGATTTTAGCAAAATACCGCGAAGATTCATCGAACCAATTAAGTGGATTGATGTC 1213  
DB 1238 GATTCGGGATTTGGATAAGTAGCGGAAAGATGGGTCTGATTGAAATGGGTTGAAGTTC 1297  
QY 1214 CAGATTGCTTTCGCTTCATCTCTGGAAGCTTTGGAGAGCGCAGAACAGATGAGTTCG 1273  
DB 1298 CTGATTGTTTCTGTTTCCATCTCTGGAATGCTTGGGAGAACACAGAAACAGATGAAATCG 1357  
QY 1274 TCGTGTAGGTCCTGTATGACTCCACAGACTCAATTTTCAACGAGTCTGACGAGAAATC 1333  
DB 1358 TTGTAATTTGTTCACTGATTCACACACACAGACTCCATTTTCAATGATGTGATGAAGAC 1417  
QY 1334 TCAAGAGTGTCTGTCTGAAATCCGCTGAATCTCAAAACCGGTGATCACTCGCCGTC 1393  
DB 1418 TAAAAAGCGTTTATCCGAAATTCGTCTCAATTTGAAACACGGGAAATCAACAAGAAAG 1477  
QY 1394 CGATCATCTCCACGAAGATCAACAGTCAACTCGAAGCAGGATGGTCAACAGAAACA 1453  
DB 1478 CCAATATCGAAACCCCGATGACAAAGTGAATTTAGAGCTGGAATGGTGAACAGAAACA 1537  
QY 1454 TGTCGCGCGCTAAACCAATTCGTTACTTGGCTTTAGCCGAGCGCTGGCCTTAAGTCT 1513  
DB 1538 AACTTGGAAAGGAAAAACACAGTATGCTTATTGGCTATTGCTGAAACCATGGCCAAAAGTTT 1597  
QY 1514 CAGGATTCGCTAAGTTGATCTCCTACTGAGAGACTTAAGAAACATCTTTAGCGGCGATA 1573  
DB 1598 CTGGTTTGCAAAAGTAGACCTGTTTACAGGTGAAGTTGAGAAATTCATTTATGGTGACA 1657  
QY 1574 ACCGTTACGGAGGAGAGCTCTGTTTCTCCCGGAG- - - - -AAGGAGGAGAGGAGAGCG 1627  
DB 1658 ACAAATATGTTGGGAGCCTCTTTTTCACCAAGAGACCCAAACACAGAGGAGGAGAGCG 1717  
QY 1628 AAGGATACATCTCTGTTTCGTTTCAGCAGAGAGACTGGAAATCGGAGTTACAGATAG 1687  
DB 1718 ATGGTTATATTTTAGCTTTCGTTTCGTTTCATGATGAGAAAGATGGACATCAGAGCTTCAAATTG 1777  
QY 1688 TTAACCGCGTTAGCTTTAGAGGTTTGAAGCAACCGGTTAAACTTCGCTCAAGGGTTCCGTACG 1747  
DB 1778 TTAATGCAATGACTTTTAAAGTTAGAGGCAACAGTGAACCTTCCATCAAGAGTTCCCTTATG 1837  
QY 1748 GATTTCACGGTACATTCATCGGAGCGCGATGANTTTCGCGAAGCAGG 1792  
DB 1838 GTTTTCATGGAAGCTTTTCATAAATGCAATGATTTGGCAAAATCAGG 1892

```
RESULT 13
AX148312
LOCUS AX148312
DEFINITION Sequence 11 from Patent EP116794.
ACCESSION AX148312
VERSION AX148312.1 GI:14347199
KEYWORDS Vigna unguiculata (cowpea)
ORGANISM Vigna unguiculata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Vigna.
REFERENCE
1. Tuchi, S., Kobayashi, M. and Shinozaki, K.
Transgenic plants carrying neoxanthin cleavage enzyme gene
Patent: EP 116794-A 11 18-JUL-2001;
Riken (JP)
FEATURES
Location/Qualifiers
1..1839
/organism="Vigna unguiculata"
/mol_type="genomic DNA"
/db_xref="taxon:3917"
1..1839
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC41197.1"
/db_xref="GI:14347200"
/translations="MPSASNTWFNATLPPFKDLPTSSNTNLLPLRKTSNTIT
CSLQTLHPKQVPTSTSTGATTTPTTKITTTTTPRETPLSDNQLPQKW
NFLQKAATLDELVEALVSHERKHPKPTADPRVOIAGNAPFVPEHAADOLPVVGK
IPKIDGVVRNGANPLVBPVAGHFFDGDGMVHAKVTNGAASVACRETFRLSQE
KSLGRPVFPAIGELHGHSGIARLLAFYARGLFLVDGSGMGVANAGLVYFNHLLA
MSDDLPHYRITPNGDLITVGRYDFNGQLNSTMIAPKLDLPVGDHLHALSYDVLQKP
YLKFFRSPGVKSPDVEIPLKEPTWMHDFAITENFVDFDQVVFKLHTEMIGSPV
VYDNKTSRFGILHKNADKANRWIDAPDFCFHLNWAWESETESEVVVIGSCMPA
DSIPNEBESLKSVLISRLNLTGKSTRPIISDAEQVNLEAGMWRNKLGRKQFPA
YLALAEWPVKVPAKVDLLSGEVKMYGEEFGLPLPNGQXEDDGYILAFVHD
EKWKSLEQIVNAQNLKLEASIKLPSRPVPGHGTPIHSKDLRKQA"
BASE COUNT 447 a 595 c 442 g 355 t
ORIGIN
Query Match 44.4%; Score 798.4; DB 6; Length 1839;
Best Local Similarity 70.4%; Pred. No. 6.7e-229;
Matches 1098; Conservative 0; Mismatches 456; Indels 6; Gaps 2;
QY 233 CCAAGAAATCAACACTAAACAGATGAATTTGTTCCAGAGAGCGCGCGCAGCGTTGG 292
DB 281 CCAACCAACCAATACCTCAAAATGGAACCTTTCTCCAGAAAGCGCTGCCACGCGCTTG 340
QY 293 ACGCGCGGAGGGTTTCCTGTGACGCCACGAGAGCTACACCCGCTTCTTAAACCGCGTG 352
DB 341 ACCTGGTCAACACGCGCTGCTGCGACGAGCGCAACACCCGCTCCCAAAACGCGCG 400
QY 353 ATCTAGTGTTCAGTATCCCGGAATTTGCTCGCGTGAATGACAGCGCGTCCGCGGTA 412
DB 401 ACCCGAGGTCCTCAATCCCGGGAACCTTCCGCGCGTCCCGAGAGTCCCGCGATCAAG 460
QY 413 ATCTTCCGCGTGTCCGAAAACCTTCCCGATTCATCAAGAGAGTGTATGTGCGCAACGAG 472
DB 461 GACTCCCGGTGTCCGAAAATCCCAATGATGATGACGCGGTGTACGTGCGCAACGTTG 520
QY 473 CTAAACCACTTCCAGAGCGGTGACAGTACCACTTCTTCAGCGGAGACGGTATGTTTC 532
DB 521 CCAATCCGCTCTACAGAGCTGTGCGCGGCAACCTTCTTCCAGCGGACCGGATGGTTC 580
QY 533 ACGCGGTCAAAATTCGAAACACGGTTTCAAGTGTAGTACGTAGTTCGCGTTTACTCAGACTAAC 592
DB 581 ACGCGGTGAGTTTACGACGCGCGCGCAGCTACGCTCGCGCTTCCACCGAGAGCGAGC 640
QY 593 GGTGTGTTCAGAAACGTCAATTTGGGTGACCGGGTTTCCCAAGCCATCGGTAGCTTC 652
```

```
641 GTCTCTCGCAGAGAGAAATCTTAGCGCGCCCGGTGTTCCGGAAGCCATCGGGAGCTCC 700
653 ACGGCCACACCGGTAATGCCCGACTCATCTATTCTACGCCAGAGCTGAGCGCGGTATAG 712
701 ACGGCCACTCCGCGCATCGCGCGGCTCCTCTCTTCTACGCGCGGTCTCTTCGGGCTCG 760
713 TCGACCGGCGACACGGAACCGGTGTAAGTAAAGCCCGGTTTGGTCTATTTCATATGGCCGT 772
761 TTGATGGGTCCCAGGCGATGGGCGTGGGAAACGCGGTCTCTCTACTTCAACACCAACC 820
773 TATTGGCTATGTCCGAGGATGATTTACCTTACCAAGTTCAGATCATCTCCCAATGGAGATT 832
821 TCTTGGCCATGTCCGAAGACGATTTACCTTACCACGTGAGAAATCACCCCTTAACGGGACT 880
833 TAAAAACCGTTGTCGGTTCGATTTTGTGAGCAATTTAGAAATCCCAAAATGATTCGCCACC 892
881 TAACACCGTGTGGCGGTACGACTTCAACGGGAGCTCACTCAACANATGATCGCCACC 940
893 CGAAAGTCGACCCCGGAATCCGGTGAATCTTTCGCTTTAAAGTACGACGCTGTTTCAAGC 952
941 CGAAACTGGACCCCGTCGACGGGACCTCCACGCGTCACTACGACGTCATTCAGAAGC 1000
953 CTTACCTAAATATCTTCCGATTCCTCACGGACGGAATTAATCACGGAGCTCGAGATTC 1012
1001 CTTACCTCAAGTACTTCCGTTTCTCCCGACGCGCTCAAGTCCCGCAGCTGGAATTC 1060
1013 AGCTTGATCAGCCAAACGATGATGACGATTCGCGATTACAGAGAACTTCGTCGTCTGATC 1072
1061 CCCTGAAGGAGCCACCATGATGACGATTTCCGCATTAACGGAATTTCTGTCGTCTGCC 1120
1073 CTGACCGCAAGTCTGTTTCAAGTCTCGGAGATGATCCGCGGTGGGTCTCCGGTGTGTT 1132
1121 CCGACCGACGAGTGGTCTTCAACTAACGAGATGATCACCGCGGCGGTCCCGCTGTGCT 1180
1133 ACGCAAGAACAAAGTCCGAGATTTCCGGATTTTAGACAAATACGCCGAAGATTCATCGA 1192
1181 ACGCAAGAACAAACCTCACGGTTTGGGATTTGCAACAAGATGCGAAGACGCGAATG 1240
1193 ACATTAAGTGAATGATGCTCCAGATTCCTCTGCTCTCTGCTTCCATCTCTGAAACGCTGGGAAG 1252
1241 CGATCGGTGATGACGCGCGGATTTGTTCTGCTTCCACCTCTGAAACGCTGGGAGG 1300
1253 AGCCAGAAACAGATGAAGTCTGCTGATGAGGTCTCTGATGACTCCACAGACTCAATTT 1312
1301 AGCCCGAAACGAGGAGGTCTGCTGATGAGTCTGATGATGATGATGATGATGATGATGAT 1360
1313 TCAACGAGTCTGACGAGAAATCTCAAGAGTCTCTGCTGAAATCGGCTGATCTCAAAA 1372
1361 TCAACGAAATCGCAGGAGAGTTTGAAGACGCTGCTCTCAGAGATGAAGGCTGAACCTGAGGA 1420
1373 CCGGTGAATCACTCGCGCTCCGATCTCTCAACGAGAGATCAACAAAGTCAACCTCGAAG 1432
1421 CCGGCAAGTCCACTCGGCGCCCATATCTCTCGACGCGG---ACAAAGTGAACCTGGAAG 1477
1433 CAGGAGTGTCAACAGAAACATGTCGCGCTGAAACCAAAATTCGCTTACTTGGCTTTAG 1492
1478 CCGGATGTGTGAACAGAAACAAAGCTCGGAAGAGAGACCAGTTCGCGTATCTGGCTCTGG 1537
1493 CCGAGCGTGGCTTAAGTCTCAGGATTCGCTAAAGTTGATCTCACTACTGGAGAGTTA 1552
1538 CCGAGCGCTGGCGCCAAAGTCTCGGCTTTGCAAAAGTTGATTTGCTGAGTGGGAGTGA 1597
1553 AGAAACATCTTTACCGCGATACCGGTACGGAAGAGAGCCTCTGTTTCTCCCGGAGAG 1612
1598 AGAAGTACATGTATGGAGAGAGAGAGTTCGGTGGGAGGCTCTGTTTCTTCTTCCC---AAGC 1654
1613 GAGGAGAGAGAGAGAGATACATCTCTGTTTCTGTTCCAGCAGCAGAGAGATCGAAAT 1672
1655 GCCAAAAGAGAGAGATGGGTATATTTGGCAATCTGTCACCGACGAGAGAGATGAAAT 1714
1673 CGGAGTTACAGATAGTTAACCGCGGTAGCTTAGAGTTGAAGCAACGGTTAACTTCGCT 1732
1715 CCGAGCTGCAGATTGTGAATGCCCAAAATTTAAAGCTCGAAGCTTCCATCAAACTCCCT 1774
```

QY 1733 CAAGGTTCCGTTACGATTTTACGGTACATTCATCGGAGCGGATGTTGGCAGCAGG 1792  
Db 1775 CTCGTTCCCTACGGTTTTCATGGAATTTTCATTTCATTCAGGATTTGAGAAACAAG 1834

RESULT 14  
BD017434 1839 bp DNA linear PAT 27-AUG-2002  
LOCUS Transgenic plant using neozanthine cleaving enzyme gene.  
DEFINITION BD017434  
ACCESSION BD017434  
VERSION BD017434.1 GI:22558610  
KEYWORDS JP 2001258579-A/6.  
SOURCE Vigna unguiculata (cowpea)  
ORGANISM Vigna unguiculata

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Vigna.

1 (bases 1 to 1839)  
Iuchi, K., Kobayashi, M. and Shinozaki, K.  
Transgenic plant using neozanthine cleaving enzyme gene  
Patent: JP 2001258579-A 6 25-SEP-2001;  
THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH  
OS Vigna unguiculata  
PN JP 2001258579-A/6  
PD 25-SEP-2001  
PF 11-JAN-2001 JP 2001003476  
PI KIYOSHI IUCHI, MASATOMO KOBAYASHI, KAZUO SHINOZAKI PC  
C12N15/09,A01H5/00,C12N5/10,C12N5/02,C12N5/00,C12N5/00 CC  
Transgenic plant using neozanthine cleaving enzyme gene PH  
Location/Qualifiers  
FT CDS (1) ..(1839) .

FEATURES  
Source 1. .1839  
/organism="Vigna unguiculata"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:3917"  
BASE COUNT 447 a 442 g 355 t  
ORIGIN

Query Match 44.4%; Score 798.4; DB 6; Length 1839;  
Best local Similarity 70.4%; Pred. No. 6,7e-229;  
Matches 1098; Conservative 0; Mismatches 456; Indels 6; Gaps 2;

QY 233 CCAAGAAATCCACACTAAACAGATGAATTTGTTCCAGAGAGCGGCGGCGAGGTTGG 292  
Db 281 CCAACCAACATTACCTCAAAATGGAATTTCTCCAGAAAGCGCTGCCACGGCTTGG 340  
QY 293 ACGGCGGAGGTTCTTTGTCAGCCACGAGAGCTACACCGCTTCTTAAACGGCTG 352  
Db 341 ACCTGTTGAAACGGGCGCTGCTCTCGACGAGCGCAACACCGCTCCCCAAACGGCGG 400  
QY 353 ATCTAGTGTTCAGATCGCGGAAATTTGCTCCGGTGAATGAACAGCCCGTCCGCGTA 412  
Db 401 ACCGAGGTTCAAATCCCGGAACTTCGCGCGGTGCGGAGCATGCGCGGATCAAG 460  
QY 413 ATCTCCGTTGTCGAAAATCTCCGATTCATCAAGAGTGTATGTGCGCAACGGAG 472  
Db 461 GACTCCCGTGTTCGAAAATCCCAATATGATGACGCGGTGTACTGTCGCAACGGTG 520  
QY 473 CTAACCACTTCAGAGCGGTGACAGGTCAACACTTCTTCGACGAGAGCGGTATGTTTC 532  
Db 521 CCAATCCGCTCTACGAGCTGTGGCGGGACCACTTCTTCGACGCGCATGCTTC 580  
QY 533 ACGCGGTCAATTCGAACACGGTTTACGTAGCTACGTTCCCGGTTTACTTCAGACTAAC 592  
Db 581 ACGCGGTGAAGTTCAACGAACGGCGCCAGCTACGCTCCCGCTTCACCGAGCGCAG 640  
QY 593 GGTTCGTCAGAACGTCATTTGGTTCGACGGGTTTCCCAAGCCATCGGTGAGCTTC 652  
Db 641 GTCTCTCGCAGGAGAAATCTCTAGGCGCGCGGTTTCCCGAAGGCCATCGGGAAGCTCC 700

653 ACGGCCACACCGGTATTGCCGACTCATGCTATTCTACGCCAGAGCTGCAGCGGTATAG 712  
701 ACGGCCACTCCGGCAATCGCGGCGCTCCTCTCTTCTACGCGCGGTCTCTTCGGGCTCG 760  
713 TCGACCGGCGACAGGAAACGGGTAGCTAAACGCCGTTTGGTCTATTTCATATGCGCGT 772  
761 TTGATGGGTCCAGGSCATGGCGTGGCGAACGCCGCTCTCGTCTACTTCAACACCAACC 820  
773 TATTGCTATGTCCGAGGATGATTTACCTTACCAAGTTCAGATTCAGATTCATCCATGGAGT 832  
821 TCTTGCCCATGTCGGAAGACGATTTACCTTACCCACGTGAGAAATCACCCCTAACGGGACT 880  
833 TAAAAACCGTTGCTCGGTTTCGATTTTATGAGCAATTAGAATCCAAATGATTTGCCCAACC 892  
881 TACCACCGTTGCGGCTTACGACTTCAACGGGCGCTCAACTCAACAAATGATCGCCCAACC 940  
893 CGAAAGTCGACCGGAAATCGGTGAATCTTTCGCTTTTAAAGTACGACGCTGTTTCAAAGC 952  
941 CGAAACTGGACCCCGCTCGACGGCGACTCCACGCGCTCAGCTACGACGCTCATTCAGAAGC 1000  
953 CTTACCTTAAATACTTCCGATTTCTACCGGAGGAACTAAATCACCGGAGCTCGAGATTC 1012  
1001 CTTACCTTAAATACTTCCGATTTCTCCCGAGCGGCTCAAGTCCCCCGAGCTGGAATCC 1060  
1013 AGCTTGATCAGCCACGATGATGATGATTCGCAATTCGCAATACAGAACTTCGTCGTCTAC 1072  
1061 CCCTGAAGGAGCCCAACCATGATGATGATTCGCCATAACGGGAAATTTGCTGCTGCTCC 1120  
1073 CTGACGAGCAAGTCGTTTCAAGCTCGCGGAGATGATCCGCGTGGGTCTCCGCTGCTTT 1132  
1121 CCGACCAAGCGGTTGTTCTTCAACTAACGGAGATGATCACCGGCGGTCTCCCGCTGCTCT 1180  
1133 ACGACAGAAACAAGGTCGCAAGATTCGGGATTTTAGACAAATACGCGGAGATTCATCGA 1192  
1181 ACGACAGAAACAACCTCACGTTTGGGATTTGCAAGAAATGCAAGAGCGCAATG 1240  
1193 ACATTAAGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1252  
1241 CGATGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1300  
1253 AGCCAGAAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1312  
1301 AGCCAGAAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1360  
1313 TCAACGAGTTCGACGAAATCTCAAGAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1372  
1361 TCAACGAAATCGAGGAGAGTTTGAAGAGCGTGTGTCAGAGATAAGGCTGAATCTGAGGA 1420  
1373 CCGGTGATCAACTCGCGCTCCGATCATCTCCAAAGAGATCAACAGATCAACCTCGAAG 1432  
1421 CCGCAAGTCCACTCGCGGCGCCCATTAATCTCCGAGCGCG---AACAGTGAACCTGGAAG 1477  
1433 CAGGATGTTCAACAGAAACATGCTCGCGCGTAAACCAAAATTCGCTTACTTGGCTTTAG 1492  
1478 CCGCATGGTGAACAGAAACAAAGCTCGGAAGGAGACCCAGTTTCGCGTATCTGCGCTCGG 1537  
1493 CCGAGCGGTGCTTAAAGTCTCAGGATTCGTAAGTTGATCTCCTACTTGGGAGAGTTA 1552  
1538 CCGAGCGCTGCCCAAGTCTCGGCTTTGGAAAGTTGATTTGCTGAGTGGGGAAGTGA 1597  
1553 AGAAACATCTTTACGGCGATAACCGTTACGAGAGAGAGCTCTGTTTCTCCCGGAGAG 1612  
1598 AGAAGTATCATGTATGGAGAGAGAAAGTTCCGTTGGGAGAGCTCTCTTCTTCTCC---AACG 1654  
1613 GAGGAGAGAGAGAGAGAGATACATCTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTT 1672  
1655 GCCAAAAAGAGAGAGATGGGTATATTTGGCATTTCTGTCGACGAGAGAGAGATGGAAT 1714  
1673 CCGGATTCAGATAGTTTAAACCGCGTTAGCTTTAGAGGTTGAGCAACGGTTTAACTCCGT 1732  
1715 CCGAGCTCGATTTGTGAATGCCCAAAATTTTAAAGTTCGAGCTTCCATCAAACTCCCT 1774  
1733 CAAAGGTTCCGTACGGATTTTACGGTATTCATTCGGAGCGGATGATTTGCGGAGAGCAGG 1792

```

Db      1775 CTGCTGTTCCACGTTTTCATGGAACTTTCATTATCCAGGATTGAGAAACAG 1834
|||||
RESULT 15
AB030293
LOCUS      2432 bp mRNA linear PLN 01-DEC-2000
DEFINITION Vigna unguiculata CPRD65 mRNA for neoxanthin cleavage enzyme,
            complete cds.
ACCESSION AB030293
VERSION    1
KEYWORDS   GI:9857289
SOURCE     neoxanthin cleavage enzyme.
ORGANISM   Vigna unguiculata (cowpea)
            Vigna unguiculata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Vigna.
REFERENCE 1 (sites)
AUTHORS   Iuchi,S., Kobayashi,M., Yamaguchi-Shinozaki,K. and Shinozaki,K.
TITLE     A stress-inducible gene for 9-cis-epoxycarotenoid dioxygenase
          involved in abscisic acid biosynthesis under water stress in
          drought-tolerant cowpea
JOURNAL   Plant Physiol. 123 (2), 553-562 (2000)
MEDLINE   20317197
PUBMED    10859185
REFERENCE 2 (bases 1 to 2432)
AUTHORS   Iuchi,S.
TITLE     Direct Submission
JOURNAL   Submitted (16-JUL-1999) Satoshi Iuchi, RIKEN, Plant Mol. Bio.;
          3-1-1 Kouyadai, Tsukuba 305-0074, Japan
          (E-mail:iuchi@rtc.riken.go.jp, Tel.81-298-36-4359)
FEATURES
            source
            1..2432
            /organism="Vigna unguiculata"
            /mol_type="mRNA"
            /db_xref="taxon:3917"
            1..2432
            /gene="CPRD65"
            126..1964
            /gene="CPRD65"
            /codon_start=1
            /product="neoxanthin cleavage enzyme"
            /protein_id="BAB11932.1"
            /db_xref="GI:9857290"
            /translations="MPSASNTWENATLPSPPKDLPTSSPTNLLPLRKTSSNTIT
            CSLOTLHPKOYQSTSTSTATTPTPKTTIATTPPRTNLSLSTNOLPKQW
            NFOKAATALDUVETALVSHERKHELPTKADPRVQIAGNFAFVPHRADQSLPVGK
            IPKIDGVYVRNAGNPLYEPVAGHFFDGGWHAVKFNGAASYACRTETQLRSQE
            KSLGRPVYPAKIGELHSGIARLLFYARGLFGLVDGSGMGVNAVGLVYFNNHLLA
            MSDDLPHYRITPGLTITVGRYDFNGQLNSTIAHPKLDVDPDGLHALSYDVQKP
            YLKYFRSPGVKSPDVEIPLKEPTWMHDFATENFVVPDQVWFKLTEMITGGSPV
            VYDKNTRSGLILHNAKADANRWIDAPDFCFHLNWAEBETBEVVVIGSMTPA
            DSIPTCEESLSKVLSEIRLNLRGTSTREPIISDABQVNLBAGVNRNKLGRKTPA
            YLALAEHPKPVSGFAYKVDLLISGEVKVMYGEKEGEPFLPLNGOKEDDGYILAFVHD
            EKEWSELQIVNAONLKEASIKLPSRPIGFHGTFIHSDLRKQA"
BASE COUNT 654 a 711 c 533 g 534 t
ORIGIN
Query Match 44.4%; Score 798.4; DB 8; Length 2432;
Best Local Similarity 70.4%; Pred. No. 7.2e-229;
Matches 1098; Conservative 0; Mismatches 456; Indels 6; Gaps 2;
QY 233 CAAAGAAATCCAACTAAACAGATGAATTTGTTCCAGAGAGCGCGCGGCGAGCGTTGG 292
Db 406 CCAACCAACCATTAACCTCAAAATAGAACTTTCTCCAGAAAGCGCGTCCACGCGCTTG 465
QY 293 ACGCGCGGAGGGTTCTCTTCAGCCACGAGAGCTACACCCGCTTCCTTAAACACGGCTG 352
Db 466 ACCTGGTCGAACGCGCTCGTCTCGACAGCGCAACACCCGCTCCCAAAACGCGGG 525
QY 353 ATCCTAGTGTTCAGATCCCGGAAATTTTGTCCGGTGAATGAACAGCCGCTCCGGCGTA 412
|||||
526 ACCGAGGGTCCAAATCGCGGGAACCTTCGCGCGGTGCGGAGCAGTCCGCGGATCAAG 585
413 ATCTTCGGTGTGCGGAAACTTCCCGATTCCATCAAGGAGTGTATGTGCGCAACGGAG 472
586 GACTCCGGTGTGCGGAAATAATCCCAATGCAATGACGGGTGTACGTGCGCAACGGTG 645
473 CTAAACCACTTCACGAGCGGTGACAGTCAACACATCTTTCGACGAGGACGGTATGGTTC 532
646 CCAATCCGCTCTACGAGCGCTGTGCGCGGCGCACCACTTCTTCGACGCGGACGGATGGTCC 705
533 ACGCGCTCAAAATCGAACACGGTTTCAGCTAGCTACGCTTTCGCCGGTTTACTTCAGACTAAAC 592
706 ACGCGCTGAAGTTCAAGAACGGCGCGCCAGCTACGCTGCCGCTCCACCGAGACGACG 765
593 GGTTTGTTCAGGACGGTCAATGGGTGACCGGTTTTTCCCAAGCAATCGGTGAGTTC 652
766 GTCTCTCGCAGGAAATCTCTAGGCGCGCGGTGTTCGGAAGGCCATCGGGAGCTCC 825
653 ACGSCACACCGGTATTGCCCGCTCATGTATTCTACGCGAGCTGCGAGCCGGTATAG 712
826 ACGCCCACTCGGCAATCGCGGGCTCTCTCTTCTACGCGCGGGTCTCTTCGGGCTCG 885
713 TCGACCCCGCACACGGAACCGGTGTAGCTAACCGCGGTTTGGTCTATTCAATGGCCGGT 772
886 TTGATGGGTCCAGGGCATGGGGTGGCGAAGCGCGGTCTCGTCTACTTCAACAACCAAC 945
773 TATTGGCTATGTCGGAGGATGATTACTTACCAAGTTCAGATCACTCCCAATGGAGATT 832
946 TCTTGGCCATGTCCGAAGACGATTTACCTACACGTGAGATCAACCCCTAACGGGACT 1005
833 TAAAAACCGTTGGTTCGATTTTGATGGACAATTTAGAAATCAACAATGATGCCCAAC 892
1006 TAAACACCGTTGGCGGTACGACTTCAACGGGAGCTCAACTCAACAATGATGCCCAAC 1065
893 CGAAGTCGACCCCGGAATCCGGTGAATCTTTCGTTTAAAGTACGAGCTGTTTCAAGC 952
1066 CGAACTGGACCCCGTCGACGCGACCTTCAACGGGCTCAAGTCCCGCGAGCTGGAATCC 1185
953 CTTACTTAAATACTTCCGATTCTCACCGGACGGAATTAATCACCGAGCTCGAGATTC 1012
1126 CTTACTCAAGTACTTCGGTTTCTCCCGCGAGCTCAAGTCCCGCGAGCTGGAATCC 1185
1013 AGCTTGATCACCACAGATGATGACGATTTCCGGATTACAGAGAACTTCGTGTCGTAC 1072
1186 CCCTGAAGAGACCCACCACTGATGACGATTTTCGCCATACAGGAGAAATTTGTCGTCGTC 1245
1073 CTGACCAACAAAGTCGTTTCAAGTTCGCGGAGATGATCCCGGTGGGTCTCCGGTGGTT 1132
1246 CCGACCAAGAGTGTGTTTCAACTAACGGAGATGATCACCGCGGGTCCCGCGTGTCT 1305
1133 ACGACAAGAACAGGTTCGAGATTCGGGATTTTAGACAATACGCCGAGATTCATCGA 1192
1306 ACGACAAGAACAAACCTCACGGTTTGGGATTTTCGCAAGAAATCGCAAGGACCGGAATG 1365
1193 ACATTAAAGTGAATGATGCTCCAGATTGCTTCTGCTTCCATCTCTGGAACGCTTGGGAAG 1252
1366 CGATCGGTGGATCGACGCGCGGATTTGTTCTGCTTCCACCTCTGGAACGCGTGGAGG 1425
1253 ACCCAGAAACAGATGAATCGTGTGATAGGGTCTGTATGACTCCACAGACTCAATTT 1312
1426 AGCCCGAAACCGAGAGGTTGTGGTGTATTGGTCTCTGCATACCCCTCGGACTCCATTT 1485
1313 TCAACGAGTCTGACAGAAATCTCAAGAGTGTCTGTCTGAAATCGCCTGAATCTCAAAA 1372
1486 TCACGAATCGAGGAGATTTGAAGAGGTGCTGTACAGATAAGGCTGAACCTGAGGA 1545
1373 CCGGTGAATCAACTCGCGTCCGATCATCTCCAAAGAAATCAACAAGTCAACCTCGAAG 1432
1546 CCGGCAAGTCCACTCGGCGCCCAATTATCTCCGACGCG--AACAAAGTGAACCTGGAAAG 1602
1433 CAGGGATGTCAACAGAACATGCTCGCGCGTAAACCAAAATTCGCTTACTTGGCTTTAG 1492
1603 CCGGCATGTGAAACAAACAAAGCTCGGAAGGACCCAGTTCGCGTATCTGGCTCTGG 1662

```

QY	1493	CGAGCCGTGGCTAAAGTCTCAGGATTCGCTAAAGTTGATCTCACTACTGGAGAAGTTA	1552
Db	1663	CGAGCCCTGGCCCAAGTCTCGGGCTTTCCGAAAGTTGATTTGCTGAGTGGGGAAGTGA	1722
QY	1553	AGAAACATCTTTACGGCGGATACCGTTACGGAGGAGCCTCTGTTTCTCCCGGAGAAG	1612
Db	1723	AGAGTACATGTATGGAGAGAGAGTTCGGTGGGAGCCTCTGTTTCTTCCC---AAGC	1779
QY	1613	GAGGAGGAAGACGAAGATACATCCTCTGTTTCGTTCCAGCAGAGAGACATGGAAT	1672
Db	1780	GCCAAAAGAGACGATGGGTATATTCTGGCATTCTGCACGACGAGAAAGAAATGGAAT	1839
QY	1673	CGAGTTACAGATAGTTAACCGGTTAGCTTAGAGTTGAAGCAACGGTTAACTCCGT	1732
Db	1840	CCAGCTGCAGATTGTGAATGCCCAAAATTTAAAGCTCGAAGTCCATCAAACTCCCT	1899
QY	1733	CAAGGTTCCGTACGGATTTACGGTACATTCATCGAGCCGATGATTTGGCGAAGCAGG	1792
Db	1900	CTCGTGTCCCTACGGTTTTCATGGAACCTTCATTCAATCCAAGGATTTGAGGAAACAAG	1959

Search completed: November 16, 2003, 17:24:03  
 Job time : 6764 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2003, 09:37:17 ; Search time 506 Seconds

(without alignments)

9602.744 Million cell updates/sec

Title: US-09-758-269-5

Perfect score: 1800

Sequence: 1 atggctttttcaaggcaac.....tggcgagcaggtcgtgtga 1800

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: N\_Geneseq\_19Jun03.\*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1800	100.0	1800	22 AAD09396	Arabidopsis thalia
2	824.6	45.8	1818	22 AAD09401	Lycopersicon escul
3	798.4	44.4	1839	22 AAD09399	Vigna unguiculata
4	738.2	41.0	1752	22 AAD09394	Arabidopsis thalia
5	606.6	33.7	1815	22 AAD09400	Zea mays neoxanthi
6	528.2	29.3	1734	22 AAD09398	Arabidopsis thalia
7	239.8	12.8	443	21 AAC56678	Eucalyptus grandis
8	204	11.3	492	22 AAC82706	Rice abscisic acid

9	190.4	10.6	393	25 ABX20484	Human GDP-mannose
10	188	10.4	372	21 AAC56695	Eucalyptus grandis
11	171.8	9.5	325	21 AAC56548	Eucalyptus grandis
12	169.6	9.4	1788	21 AAC42989	Arabidopsis thalia
13	169.6	9.4	1788	22 AAD09395	Arabidopsis thalia
14	169.6	9.4	1788	24 ABZ13639	Arabidopsis thalia
15	169.2	9.4	1950	22 AAF77206	CDNA encoding sunf
16	149.4	8.3	491	21 AAC57157	Pinus radiata tran
17	143.4	8.0	386	21 AAC57162	Pinus radiata tran
18	140.4	7.8	326	21 AAC57167	Pinus radiata tran
19	139.8	7.8	1617	22 AAD09397	Arabidopsis thalia
20	139.8	7.8	1617	22 AAD09402	Arabidopsis thalia
21	139.8	7.8	1777	21 AAC36083	Arabidopsis thalia
22	137.2	7.6	329	21 AAC57165	Pinus radiata tran
23	121.6	6.8	398	21 AAC57145	Pinus radiata tran
24	118	6.6	340	21 AAC56520	Eucalyptus grandis
25	113	6.3	238	21 AAC57169	Pinus radiata tran
26	109.6	6.1	246	21 AAC57168	Pinus radiata tran
27	106.6	5.9	412	21 AAC57164	Pinus radiata tran
28	93	5.2	320	25 ABX20813	Human GDP-mannose
29	89	4.9	190	21 AAC57163	Pinus radiata tran
30	78.6	4.4	303	21 AAC57125	Pinus radiata tran
31	76.8	4.3	308	21 AAC57142	Pinus radiata tran
32	74.8	4.2	395	21 AAC57117	Pinus radiata tran
33	72	4.0	470	24 ABLS7537	Apoptosis inhibito
34	67.8	3.8	285	21 AAC56908	Pinus radiata tran
35	65.8	3.7	567	21 AAC56546	Eucalyptus grandis
36	64.6	3.6	457	21 AAC56910	Pinus radiata tran
37	63	3.5	445	21 AAC56948	Pinus radiata tran
38	62.4	3.5	436	21 AAC57124	Pinus radiata tran
39	60.8	3.4	367	24 ABQ85532	Arabidopsis thalia
40	58.2	3.2	304	21 AAC56973	Pinus radiata tran
41	56.6	3.1	419	21 AAC56527	Eucalyptus grandis
42	56.6	3.1	447	21 AAC56480	Eucalyptus grandis
43	56.6	3.1	508	21 AAC56611	Eucalyptus grandis
44	55.6	3.1	4403765	22 AA193683	Mycobacterium tube
45	55.6	3.1	4411529	22 AA193682	Mycobacterium tube

ALIGNMENTS

RESULT 1

AAD09396

ID AAD09396 standard; cDNA; 1800 BP.

XX AAD09396;

AC AAD09396;

DT 10-SEP-2001 (first entry)

XX Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED3 cDNA.

DE Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED3 cDNA.

KW Neoxanthin cleavage enzyme; AtNCED3; abscisic acid; ABA; herbicide;  
KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;  
KW plant growth protectant; ss.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

XX Key

FT CDS

FT 1..1800

FT /tag= a

FT /product= "Arabidopsis thaliana AtNCED3 protein"

XX EP1116794-A2.

XX 18-JUL-2001.

XX 11-JAN-2001; 2001EP-0300218.

XX 13-JAN-2000; 2000JP-0010056.

XX 11-JAN-2001; 2001JP-0003476.

XX (RIKE ) RIKEN KK.

XX Tuchi S, Kobayashi M, Shinozaki K;  
XX WPI; 2001-400081/43.  
XX P-PSDB; AAB04784.  
XX  
PT A DNA encoding a protein with a neoxanthin cleavage activity for  
PT producing transgenic plants with improved or decreased stress tolerance  
XX  
XX Claim 3; Page 32-36; 101pp; English.  
XX  
XX The invention relates to neoxanthin cleavage enzymes and their  
XX corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key  
XX role in endogenous abscisic acid (ABA) biosynthesis under drought stress.  
XX Neoxanthin cleavage enzyme is used for improving stress tolerance in a  
XX plant when expressed in a plant cell. The invention also relates to  
XX methods for increasing or decreasing stress tolerance in a plant by  
XX introducing the DNA into the plant, and a transgenic plant into which a  
XX neoxanthin cleavage enzyme is introduced. The improvement of stress  
XX tolerance in plants is useful, for example, in plant breeding. Neoxanthin  
XX cleavage enzyme genes are useful for producing transgenic plants. An arid  
XX land can be improved by growing transformant weed for several years and  
XX then removing the weed, by specifically lowering stress tolerance in the  
XX weed by inducing an inducible promoter. The present cDNA sequence encodes  
XX Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCE3 protein.  
XX The AtNCE3 cDNA is obtained from an Arabidopsis plant-derived cDNA  
XX library using a cDNA of the CPRD65 (Cowpea Responsive to Dehydration)  
XX gene isolated from cowpea plant as a probe.  
XX  
XX Sequence 1800 BP; 458 A; 464 C; 439 G; 439 T; 0 other;  
XX  
Query Match 100.0%; Score 1800; DB 22; Length 1800;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1800; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 ATGGCTCTTTTTCACGGCAACGGCTGCGGTTTCTGGAGATGGCTTGGTGGCAATCATACT 60  
1 ATGGCTCTTTTTCACGGCAACGGCTGCGGTTTCTGGAGATGGCTTGGTGGCAATCATACT 60  
61 CAGCGCCGATATTCGCTTCTCAAGCTCCGACTTGGATTGATTTGCTGCTCTTACCTATG 120  
61 CAGCGCCGATATTCGCTTCTCAAGCTCCGACTTGGATTGATTTGCTGCTCTTACCTATG 120  
121 GCCAGTCGTGTACACAGTAAGTCAATGTTTTCATCTCGGTTTACACTCTCCAGGCTCTT 180  
121 GCCAGTCGTGTACACAGTAAGTCAATGTTTTCATCTCGGTTTACACTCTCCAGGCTCTT 180  
181 CATTTCCCTTAGCAATCATCAACTCTCCGCCATTGTTGTAAGCCCAAGCAAGAA 240  
181 CATTTCCCTTAGCAATCATCAACTCTCCGCCATTGTTGTAAGCCCAAGCAAGAA 240  
241 TCCAAACACTAAACAGATGAATTTGTTCCAGAGAGCGCGGCGAGCGGTTGGACGGCGG 300  
241 TCCAAACACTAAACAGATGAATTTGTTCCAGAGAGCGCGGCGAGCGGTTGGACGGCGG 300  
301 GAGGTTTCCTTGTGACGCGAGAGAGTACACCGGTTCTTAAACCGGCTGATCTAGT 360  
301 GAGGTTTCCTTGTGACGCGAGAGAGTACACCGGTTCTTAAACCGGCTGATCTAGT 360  
361 GTTCAGATCGCGGAAATTTTGTCTCGGTTGAATGAACAGCCGCTCCGCGGCTAATCTTCG 420  
361 GTTCAGATCGCGGAAATTTTGTCTCGGTTGAATGAACAGCCGCTCCGCGGCTAATCTTCG 420  
421 GTGGTCGGAACACTTCCCGATTCCATCAAGAGTGTATGTCGACGAGGCTAACCCA 480  
421 GTGGTCGGAACACTTCCCGATTCCATCAAGAGTGTATGTCGACGAGGCTAACCCA 480  
481 CTTACGAGCGCGGTGACAGGTACCACTTCTTCGACGAGACGGTATGTTTACGCGGTC 540  
481 CTTACGAGCGCGGTGACAGGTACCACTTCTTCGACGAGACGGTATGTTTACGCGGTC 540  
541 AATTCGAACAGGTTACGTAGTACGCTTTCGCGGTTTACTCAGACTAACCGGTTTGT 600

541 AAATTCGAACAGGTTTACGTAGTACGCTTTCGCGGTTTACTCAGACTAACCGGTTTGT 600  
601 CAGGAACGTCAATTTGGGTGACCGGTTTCCCAAGACCATGGTGGTTCACGGCCAC 660  
601 CAGGAACGTCAATTTGGGTGACCGGTTTCCCAAGACCATGGTGGTTCACGGCCAC 660  
661 ACCGGTATTCGCGGACTCATGCTATTCTACGCCAGAGCTGCAGCCGGTATAGTCGACCGG 720  
721 GCACACGAACCGGTTAGTAAACGCGGTTTGGTCTATTCTTCAATGGCCGGTTATTGGCT 780  
721 GCACACGAACCGGTTAGTAAACGCGGTTTGGTCTATTCTTCAATGGCCGGTTATTGGCT 780  
781 ATGTGCGAGGATGATTTACCTTACCAAGTTAGATCATCTCCCAATGGAGATTAAAGAAC 840  
781 ATGTGCGAGGATGATTTACCTTACCAAGTTAGATCATCTCCCAATGGAGATTAAAGAAC 840  
841 GTTGGTTCGGTTCGATTTTGGTGGACCAATTAGAATCCACATGATTCGCCACCGCAAGTC 900  
841 GTTGGTTCGGTTCGATTTTGGTGGACCAATTAGAATCCACATGATTCGCCACCGCAAGTC 900  
901 GACCGGAATTCGGGTGAACTCTTCGGCTTTAAGTACGAGCTGCTTTCAAGGCTTACCTA 960  
901 GACCGGAATTCGGGTGAACTCTTCGGCTTTAAGTACGAGCTGCTTTCAAGGCTTACCTA 960  
961 AAATACTTCCGATTTCTCACCGGACCGAATCTAAATCAACCGAGCTCGAGATTTCAGCTT 1020  
961 AAATACTTCCGATTTCTCACCGGACCGAATCTAAATCAACCGAGCTCGAGATTTCAGCTT 1020  
1021 CAGCAACGATGATGCAGATTTTCGGGATTCAGAGATTCAGAGAACTTCGTCGCTACCTGAC 1080  
1021 CAGCAACGATGATGCAGATTTTCGGGATTCAGAGATTCAGAGAACTTCGTCGCTACCTGAC 1080  
1081 CAAGTCGTTTCAAGCTTCGCGGAGATGATCCGCGGTGGGTCTCCGGTGGTTCACGACAA 1140  
1081 CAAGTCGTTTCAAGCTTCGCGGAGATGATCCGCGGTGGGTCTCCGGTGGTTCACGACAA 1140  
1141 AACAGGTCGCAAGATTCGGGATTTTAGACAAATACGCCGAAGATTCATCGAACATTAA 1200  
1141 AACAGGTCGCAAGATTCGGGATTTTAGACAAATACGCCGAAGATTCATCGAACATTAA 1200  
1201 TGGATTGATGCTCAGATTCGTTCTCTGCTTCCATCTCTGAAACGCTTGGGAAGAGCCAG 1260  
1201 TGGATTGATGCTCAGATTCGTTCTCTGCTTCCATCTCTGAAACGCTTGGGAAGAGCCAG 1260  
1261 ACAGATGAAGTCGTCGATAGGTCCTGTATGACTCCACGACTCAATTTTCAACGAG 1320  
1261 ACAGATGAAGTCGTCGATAGGTCCTGTATGACTCCACGACTCAATTTTCAACGAG 1320  
1321 TCTCAGCAGAACTCTCAAGAGTGTCTGTCTGAAATCCGCTGAATCTCAAAACCGGTGAA 1380  
1321 TCTCAGCAGAACTCTCAAGAGTGTCTGTCTGAAATCCGCTGAATCTCAAAACCGGTGAA 1380  
1391 TCRACTCCGCTCGATCTCCAGGAGATCAACAGTCAACCTCGAAGCGGATG 1440  
1391 TCRACTCCGCTCGATCTCCAGGAGATCAACAGTCAACCTCGAAGCGGATG 1440  
1441 GTCAACGAAACATGTCGCGCGTAAACCAATTCGCTTACTTGGCTTTAGCGGCGG 1500  
1441 GTCAACGAAACATGTCGCGCGTAAACCAATTCGCTTACTTGGCTTTAGCGGCGG 1500  
1501 TGGCTTAAGTCTCAGGATTCGATTAAGTTGATCTCCTACTTGGAGAGTTAAGAACT 1560  
1501 TGGCTTAAGTCTCAGGATTCGATTAAGTTGATCTCCTACTTGGAGAGTTAAGAACT 1560  
1561 CTTTACGCGATACCGTTACGAGGAGGCTCTGTTTCTCCCGGAGAGGAGAG 1620  
1561 CTTTACGCGATACCGTTACGAGGAGGCTCTGTTTCTCCCGGAGAGGAGAG 1620  
1621 GACGAGGAGATACCTCTCTCTGTTTCAGGAGAGGAGATGGAATTCGAGTTA 1680



Db 1621 GAAGACGAAGGATACATCTCTGTTTCCTTCACGACGAGAAAGACATCGAATCGGAGTTA 1680  
 Qy 1681 CAGATAGTAAACCGCGTTAGCTTAGAGTTGAAGCAACGGTTAAACTTCCGTCAAGGGTT 1740  
 Db 1681 CAGATAGTAAACCGCGTTAGCTTAGAGTTGAAGCAACGGTTAAACTTCCGTCAAGGGTT 1740  
 Qy 1741 CGGTACGAGTTTACGGTACATTCATTCGAGCCGAGTATTTGGCGAAGCAGGTCGTGTGA 1800  
 Db 1741 CGGTACGAGTTTACGGTACATTCATTCGAGCCGAGTATTTGGCGAAGCAGGTCGTGTGA 1800

## RESULT 2

AAD09401

ID AAD09401 standard; cDNA; 1818 BP.

XX AC

XX AAD09401;

XX DT 10-SEP-2001 (first entry)

XX DE

Lycopersicon esculentum neoxanthin cleavage enzyme, LeNCED1 cDNA.

XX KW

Tomato; neoxanthin cleavage enzyme; LeNCED1; abscisic acid; ABA;

XX KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;

XX KW plant growth protectant; herbicide; ss.

XX OS

Lycopersicon esculentum.

XX FH

Key Location/Qualifiers

XX FT 1..1818

XX FT CDS

XX FT /\*tag= a

XX FT /product= "Lycopersicon esculentum LeNCED1 protein"

XX PN EF1116794-A2.

XX PD 18-JUL-2001.

XX PF 11-JAN-2001; 2001BP-0300218.

XX PR 13-JAN-2000; 2000JP-0010056.

XX PR 11-JAN-2001; 2001JP-0003476.

XX XX (RIKE ) RIKEN KK.

XX PI Tuchi S, Kobayashi M, Shinozaki K;

XX DR WPI; 2001-400081/43.

XX DR P-PSDB; ABE04789.

XX XX A DNA encoding a protein with a neoxanthin cleavage activity for

XX PT producing transgenic plants with improved or decreased stress tolerance

XX PS Claim 3; Page 67-71; 101pp; English.

XX CC The invention relates to neoxanthin cleavage enzymes and their

XX CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key

XX CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.

XX CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a

XX CC plant when expressed in a plant cell. The invention also relates to

XX CC methods for increasing or decreasing stress tolerance in a plant by

XX CC introducing the DNA into the plant, and a transgenic plant into which a

XX CC neoxanthin cleavage enzyme is introduced. The improvement of stress

XX CC tolerance in plants is useful, for example in plant breeding. Neoxanthin

XX CC cleavage enzyme genes are useful for producing transgenic plants. An arid

XX CC land can be improved by growing transformant weed for several years and

XX CC then removing the weed by specifically lowering stress tolerance in the

XX CC weed by inducing an inducible promoter. The present cDNA sequence encodes

XX CC Lycopersicon esculentum neoxanthin cleavage enzyme, LeNCED1 protein

XX CC related to the invention.

XX SQ Sequence 1818 BP; 569 A; 383 C; 369 G; 497 T; 0 other;

XX Query Match 45.8%; Score 824.6; DB 22; Length 1818;

Best Local Similarity 69.6%; Pred. No. 4.8e-257;  
 Matches 1134; Conservative 0; Mismatches 489; Indels 6; Gaps 1;

Qy 170 CTCACGCTCTTCATTTCCCTTAAGCAATCATCAAACTCTCCGCCATTTGTTGTTAAAGCCA 229  
 Db 185 CTTCAATTTATCAACACCAAGAAATATACATTTACACCCAAACAAAGAAACAACA 244  
 Qy 230 AAGCCAAAGAAATCCAACTAAACAGATGAATTTGTTCCAGAGAGGGGGGGGAGCGCT 289  
 Db 245 ACTCTCTCTCTCTTCAACTTCCAAAGTGAATTTAGTGCAGAAAGCAGCAATGGCTT 304  
 Qy 290 TGGACCGCGGGAGGGTTTCTTGTCTCAGCCACGAGAAAGCTACACCGCTTCTCTAAAACGG 349  
 Db 305 TAGATGCTGTAGAAAGTGTCTTAACTAAACATGAACCTGAAACACCTTTGCGGAAACAG 364  
 Qy 350 CTGATCCTAGTGTTCAGATCGCGGAAATTTTGTCTCCGGTGAATGAACAGCCGCTCGGC 409  
 Db 365 CCGACCCAGAGTCCAGATTTCTGGGAATTTTGTCTCCGGTACCGGAAATTCAGCTCTGTC 424  
 Qy 410 GTAACTTTCGGTGTCCGGAACACTTCCCGATTCCATCAAAAGAGTGTATGCGCAACG 469  
 Db 425 AATCTCTTCCGGTACCGGAAATTAACCAATGTTTCAAGGCTTTACGTTTCAAGAGC 484  
 Qy 470 GAGCTAACCCACTTTCAGAGCCGGTGACAGGTCAACACTTTCTTCAGCGAGACGGTATGG 529  
 Db 485 GAGCTAACCCCTCTTTTGAACCAACCGCCGACACCAATTTCTTCGACGCGACGGTATGG 544  
 Qy 530 TTCACGCGCTCAAAATTCGACACGGTTCAGCTAGCTACGCTTCCGGTCTTACTCAGACTA 589  
 Db 545 TTCACGCGCTCAAAATTCGACACGGTTCAGCTAGCTACGCTTCCGGTCTTACTCAGACTA 604  
 Qy 590 ACCGGTTTGTTCAGGAACGTCATTTGGGTGCGACCGGTTTTCCCAAGGCAATCGGTGAGC 649  
 Db 605 AGAGGCTTGTTCAGGAACGTCATTTGGGTGCGACCGGTTTTCCCAAGGCAATCGGTGAGT 664  
 Qy 650 TTCACGCGCTCAAAATTCGACACGGTTCAGCTAGCTACGCTTCCGGTCTTACTCAGACTA 709  
 Db 665 TACATGCTCACTCTGGAATTCGAAAGCTTATGCTGTTTACGCTCGTGGGCTTCTCGGAC 724  
 Qy 710 TAGTCGACCGCGCACACGGAACCGGTGTAGCTAAACCGCGTTCGTTTCAATGAGCC 769  
 Db 725 TTGTTGATCACAGTAAAGGAACTGGTGTTCGAAACCGCGGTTTGTCTTATTCATAACC 784  
 Qy 770 GGTATTGGCTATGTCGGAGGATGATTTACCTTACCAAGTTACAGTCACTCCCAATGGAG 829  
 Db 785 GATTACTTGTCTGTAAGATGATTTGCTTTACCATGTAAAGGTAAACCCACCGCGG 844  
 Qy 830 ATTTAAACCGTTCGTTTCGTTTCGATTCGCAATTTAGAAATCCAAATGATTTGATGCC 889  
 Db 845 ATCTTAAACAGAGGTTCGATTCGATTCGACGCGCAGCTAAATCCACATGATAGCTC 904  
 Qy 890 ACCGAAAGTCGACCCCGGAATCCGGTGAACCTTTTCGCTTTTAAAGTACGAGCTGTTCAA 949  
 Db 905 ACCGAAAGTCGACCCCGGAATCCGGTGAACCTTTTCGCTTTTAAAGTACGAGCTGTTCAA 964  
 Qy 950 AGCTTACCTAAATATCTTCGATTTCTCAGCGGACGGAATCAATCAGCGAGCTCGAGA 1009  
 Db 965 AGCATACCTCAAGTACTTCAAGTTTCAAAAATGGGAAATCAATGATGTTGAA 1024  
 Qy 1010 TTCAGCTTGATCAGCCCAACGATGATGACGATTCGCGATTACAGAGAACTTCGTCTCG 1069  
 Db 1025 TTCAGTTGAAGCCCAACCAATGATGATGATTTGCAATTTACTGAGAACTTCGTCTGCA 1084  
 Qy 1070 TACTGACGAGCAAGTCTGTTTCAAGTTCGCGAGATGATCCCGGTGGTCTCCGGTGG 1129  
 Db 1085 TTCTGTATCAACAGTCTGTTTCAAGTGTCTGAAATGATCCCGTGGAGGTTACCGGTGG 1144  
 Qy 1130 TTTACGCAAGAAACAAGTTCGCAAGATTTCCGGATTTTAGACAAATAGCCGAAGATTCT 1189  
 Db 1145 TTTACGCAAGAAACAAGTTCGCAAGATTTCCGGATTTTAGACAAATAGCCGAAGATTCT 1204  
 Qy 1190 CGAATTAAGTATGATGCTCCAGATTTGCTTCTGCTTCCATCTCTGAAACGCTTGG 1249



821 TCCTGGCCATGTCGGAAGACGATTTACCTTACCAGTGAAGATCACCCCTAACGGCGACT 880  
833 TAAACACGTTGGTTCGATTTTGTATGACAAATAGATTCACAAATGATTCGCCACC 892  
861 TAAACACGTTGGTTCGATTTTGTATGACAAATAGATTCACAAATGATTCGCCACC 940  
893 CGAAAGTCGACCCCGGAATTCGGTGAATCTTTCGTTTAAAGTACGATCGCTTTCAAGC 952  
941 CGAAACTGGAACCCGTCGACGCGACCTCCACGCGCTCAGCTACGATCTATTGAGAGC 1000  
953 CTTACTTAATATCTTCCGATTTCTCACCGGACGGAATTAATTCACCGAGCTCGAGATTC 1012  
1001 CTTACTTAATATCTTCCGATTTCTCACCGGACGCGGTCAAGTCCCGGACGTGGAATTC 1060  
1013 AGCTTGAATCAGCAACGATGATGACGATTTGCGGATTTACGAGAACTTCGTCGTCGAC 1072  
1061 CCCTGAAGGAGCCCAACCTGATGACGATTTCCGCAATAACGAGAAATTTCTGTCGTCGC 1120  
1073 CTGACGAGCAAGTCTGTTTCAAGCTGCGGAGATGATCGCGGTGGGTCTCGGTGGTT 1132  
1121 CCGACGAGCAGTGTCTTCAAACTAAACGAGATGATCACCGCGGGTCCCGGTGGTCT 1180  
1133 ACGACAAGAACAAAGTCGCAAGATTCGGGATTTTGAACAATATACCGCGAAGATTCATCGA 1192  
1181 ACGACAAGAACAAACCTCAAGTTTGGATTTCTGCAAGAATTCGAAGGACGCGAATG 1240  
1193 ACATTAAGTGAATGATGCTCAGATTTGCTTCTGCTTCCATCTCTGGAACGCTTGGGAG 1252  
1241 CGATGCGTGGATCGACGCGCGGATTTCTGCTTCCATCTCTGGAACGCTTGGGAG 1300  
1253 AGCCAGAAACAGATGAAGTCGTCGATAGGTCCTGTATGACTCCACCACTCAATTT 1312  
1301 AGCCAGAAACAGAGAGTGTGTTGATGATTTGGTCTCTGATGACCTCGGACTCCATTT 1360  
1313 TCAACGAGTCTGACGAGAACTCAAGAGTGTCTGTCTGAAATCGCCCTGAATCTCAAA 1372  
1361 TCAACGAAATCGAGGAGAGTTTGAAGAGCGTCTGTCTGAGATAAGGTGAATTTGAGA 1420  
1373 CGGTGAATCACTCCGCTCGATCTATCTCAACGAGATCAACAAGTCACTCTCGAAG 1432  
1421 CCGGCAAGTCCATCGCGCGCCCATTTATCTCGACGCCG---AACAGTGAACCTGGAG 1477  
1433 CAGGATGTTCAACAGAAACATGCTCGCGCGTAACAAACCAAAATCGCTTACTTGGCTTAG 1492  
1478 CCGGATGTTGAACAGAAACAGCTCGGAAGGAGACCCAGTTCGGTATCTGTGCTCTGG 1537  
1493 CCGAGCGGTGGCTTAAGTCTAGGATTCGCTAAAGTTGATCTCACTACTGAGAGTTA 1552  
1538 CCGAGCGGTGGCTTAAGTCTAGGATTCGCTAAAGTTGATTTGCTGAGTGGAGAGTA 1597  
1553 AGAAACATCTTTACGCGGATACCGTTACGAGGAGAGCTCTGTTTCTCCCGGAGAG 1612  
1598 AGAAGTACATGTATGAGAGAGAGAGTTTCGGTGGGAGGCTCTGTTTCTTCCG---AACG 1654  
1613 GAGGAGGAGGAGGAGGAGATACATCTCTGTTTCTGTTTACGAGGAGAGAGATGGAAT 1672  
1655 GCCAAAGGAGGAGGAGTGGGTATATTCTGCAATCTGTCGACGAGGAGGAGGATGGAAT 1714  
1673 CCGAGTTTACAGATAGTTAAACCGGTGTAGCTTGAAGCAACGGTTTAACTTCCGT 1732  
1715 CCGAGTCTGAGATGTGATGCCCCAAATTTAAGCTCGAAGCTTCCATCAAACTCCCT 1774  
1733 CAAGGTTCCGTACGAGATTTACGAGTACATTCATCGGAGCGGATGATTTGGCGAAGCAG 1792  
1775 CTCGTGTTCCCTACGGTTTTCATGGAATTTTCTTCAATCCAAAGGATTTGAGGAACAAG 1834

RESULT 4  
AAD09394 standard; cDNA; 1752 BP.  
XX  
AC  
AAD09394;  
XX

10-SEP-2001 (first entry)  
XX Arabidopsis thaliana neoxanthin cleavage enzyme; AtNCED1 cDNA.  
XX Neoxanthin cleavage enzyme; AtNCED1; abscisic acid; ABA; herbicide;  
KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;  
KW plant growth protectant; ss.  
XX Arabidopsis thaliana.  
XX  
FH Key Location/Qualifiers  
CDS 1..1752  
FT /tag= a  
FT /product= "Arabidopsis thaliana AtNCED1 protein"  
FT  
PN EF116794-A2.  
XX  
PD 18-JUL-2001.  
XX  
XX 11-JAN-2001; 2001EP-0300218.  
PF  
XX 13-JAN-2000; 2000JP-0010056.  
PR  
PR 11-JAN-2001; 2001JP-0003476.  
XX  
XX (RIKE ) RIKEN KK.  
XX  
XX Iuchi S, Kobayashi M, Shinozaki K;  
PI WPI; 2001-400081/43.  
XX P-PSDB; AAE04782.  
DR  
XX A DNA encoding a protein with a neoxanthin cleavage activity for  
PT producing transgenic plants with improved or decreased stress tolerance  
PT  
XX  
XX Claim 3; Page 18-22; 101pp; English.  
XX  
XX The invention relates to neoxanthin cleavage enzymes and their  
CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key  
CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.  
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a  
CC plant when expressed in a plant cell. The invention also relates to  
CC methods for increasing or decreasing stress tolerance in a plant by  
CC introducing the DNA into the plant, and a transgenic plant into which a  
CC neoxanthin cleavage enzyme is introduced. The improvement of stress  
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin  
CC cleavage enzyme genes are useful for producing transgenic plants. An arid  
CC land can be improved by growing transgenic plant weed for several years and  
CC then removing the weed by specifically lowering stress tolerance in the  
CC weed by inducing an inducible promoter. The present cDNA sequence encodes  
CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED1 protein.  
CC The AtNCED1 cDNA is obtained from an Arabidopsis plant-derived cDNA  
CC library using a cDNA of the CPRD65 (Cowpea Responsive to Dehydration)  
CC gene isolated from cowpea plant as a probe.  
XX  
XX Sequence 1752 BP; 484 A; 396 C; 428 G; 444 T; 0 other;  
SQ

Query Match 41.0%; Score 738.2; DB 22; Length 1752;  
Best Local Similarity 68.2%; Pred. No. 6.2e-229;  
Matches 1063; Conservative 0; Mismatches 478; Indels 18; Gaps 2;  
QY 244 AACACTAACAGATGAATTTGTTCCAGAGAGCGCGCGGAGCGTTGGACGCGCGAG 303  
Db 193 AATCCTCTCCGCTAAACATCTTCCAGAAAGCGCGGATTCGATCGACGCGCTGAG 252  
QY 304 GGTTCCTTGTGACGACGAGAACTACACCGCTTCTTAAACGCTGATCTTAGTGT 363  
Db 253 COTGCAATTAATCTCACACGAGCAAGATTCTCCATTTCCCAAAACCGCTGATCCACGTT 312  
QY 364 CAGATCGCGGAAATTTTCTCCGTTGAATGAACAGCGCGCTCCGCGCTTAATCTTCCG 423  
Db 313 CAGATTGCGGGAATTTATTCCTCCGTCAGCGAATCTTCCGTCGCGGAAACCTCACCGTC 372

QY 424 GTCCGAAACCTCCCGATTCATCAAAAGGAGTGATGTGCGCAACGAGGCTAACCCACTT 483  
DB 373 GAAGGAACAATCCCTGACTGCTGCAATGACGGTGTATATATCGGTAAACGCGCGAATCCGATG 432  
QY 484 CACGAGCGGTCGACAGGTCACACCTTCTCCAGCGAGACGGTATGTTTCACGCGCTCAAA 543  
DB 433 TTTGAGCCAAACAGCTGGCGCCACCTTATTCGACGGAGACGGATGTTTCAGCGAGTTAA 492  
QY 544 TTCGAACACGGTTACGCTAGCTAGCTAGCTTCCCGGTTTATCTCAGACTAACCGGTTTGTTCAG 603  
DB 493 ATACCAACCGGTTACGCTAGCTAGCTAGCTAGCTTCCCGGTTTATCAAAACCGAGAGATTTGGTTCAG 552  
QY 604 GAACGCTCAATTTGGGTTCGACCGGTTTCCCAAGCCATCGGTGAGCTTTCACGCGCCACACC 563  
DB 553 GAAAAACGATTTGGGTTCGACCGGTTTCCGAAACCAATCGGCGAGCTTTCAGCGTCACTCG 612  
QY 664 GGTATTTGCCGACTCATGCTATTTCTACCCAGAGCTGCGCGGTATAGTCGACCGCGCA 723  
DB 613 GGAATCGCAGCTTTGATGCTGTTTACGCGACGTGGGCTTTGTGCTCTGATCAACAACAA 672  
QY 724 CACGGAACCGGTGTAGCTTAACCGCGGTTTGTCTATTTCAATGCGCCGTTTATTTGGCTATG 783  
DB 673 AACCGCGTCGAGTAGCAAAACCGCGTTTGTGTTTACCTTAAATACCGGCTTTTAGCTATG 732  
QY 784 TCGGAGGATGATTTACCTTACCAAGTTCAGATCACTCCCAATGGAGATTTAAACCGTT 843  
DB 733 TCAGAAAGACGATTTACCGGTACCAATTAATAATTAATCAACCGCGGATCTCCAAACCGGT 792  
QY 844 GGTCCGTTTCGATTTTGTAGTGGCAATTAAGATCCCAATGATGTCGCCACCGGAAGTCGAC 903  
DB 793 GGACGTTACGATTTTCGACGGT-CAGTTAAATCCGCAATGATGCTACCCCGAACTGGAC 852  
QY 904 CCGGAATCCGTTGACTCTTCGCTTTAGCTTACGACGCTGCTTCAAGCGCTTACCTTAA 963  
DB 853 CCGGTTACGAGGAGCTTTCAGCGGTTAAGCTACGACGCTGTTAAGAACCTTACCTTGA 912  
QY 964 TACTTCCGATCTCACCGGACGCACTAAATCACCGGACGTCGAGATTCAGCTTGATCAG 1023  
DB 913 TACTTCCGATCTCACCGGACGCGGTTAAATCCGCGGATTTGAGATCCCGCTCGAACT 972  
QY 1024 CCAAGCATGATGACGATTTCCGATTTACAGAGACTTCGTCGCTGACCTGACCGACAA 1083  
DB 973 CCGACGATGATTCAGATTTGCTTAAACGGAGATTTTGGTGATTCCTGATCAACAA 1032  
QY 1084 GTCCGTTTCAAGCTCCCGAGATGATCCGCGGTGCTCCGCGGTTTACGACAAAGAC 1143  
DB 1033 GTCCGTTTCAAGCTCCCGAGATGATTTCCGCGTAAATCTCCGCTGTTGTTTTCGACGGAGAA 1092  
QY 1144 AAGGTCGAAGATTCGGGATTTTAGACAAATACCGGAGATTCATCGAATTAAGTGG 1203  
DB 1093 AAGGTTCCGATTCGGGATTAAGCCAGGACGCGACAGAGCTTCTAGATTAATCTCG 1152  
QY 1204 ATGATGCTCCAGATGCTTCTGCTTCAATCTGGAACGCTTGGGAGAGCGAGAAACA 1263  
DB 1153 GTGAATCTCCCGGAGACGCTTCTGTTTTCATCTCTGGAATGATGGGAATCCCGGAGAGC 1212  
QY 1264 GATGAAGTCTGCTGATAGGCTTCGTATGATCTCCAGACTCAATTTTCAACGAGTCT 1323  
DB 1213 GAGGAGATGCTGATCGGATCGTGTATGTCGCGCGGATTCATCTTCAACGAGAGA 1272  
QY 1324 GACGAGATCTCAAGATGCTCTGCTGAAATCCGCTGAAATCTCAAAACCGGTGAATCA 1383  
DB 1273 GACGAGATCTCAAGATGCTCTGCTGAAATCCGCTGAAATCTCAAAACCGGTGAATCA 1332  
QY 1384 ACTCGCGCTCCGATCATCTCCAAAGAGATCAACAACTCAACCTCGAGGAGGATGGTC 1443  
DB 1333 ACGGCTCGCTGCTGCTGCTTAAACGAGAT-----GTAAATTTAGAGATTTGGTATGTT 1386  
QY 1444 AACAGAAACATGCTCGCGCGGTAACCAAAATTCGTTACTTGGCTTTAGCGGACCGGTG 1503  
DB 1387 AACCGGAACCGGTAGGAGAAACCCCGGTTTCGCTTTTGGCTATGCTTATCTCTG 1446  
QY 1504 CCTAAAGTCTCAGGATTCGTTAAAGTTGATCTCACTACTGAGAGATTAAGAAACATCTT 1563

DB 1447 CCAAAAGTTTCCGGTTTCGCTAAGSTCGATCTTTGACCGGTGAGATGAAAAATATATT 1506  
QY 1564 TACGGCGATTAACCGTTTACGAGGAGAGCTCTCTGTTTCTCCCGG-----AGAA 1611  
DB 1507 TACGGCGGTGAGAAATATGCGCGCAACCGTTTCTTCTGCGCGCAACTCCGGTAAACGC 1566  
QY 1612 GGAGGAGAGAGACGACGAGATACATCTCTCTCTTCTGTTTCAGCGAGAGACATGGAAA 1671  
DB 1567 GAAGAAATGAAGATGACCGTTATATTTTGTACGTTTCATGCGAAGAAACAAGACA 1626  
QY 1672 TCGAGTTACAGATAGTTAACCGCTTACGCTTACGAGTTTGAAGCAACCGTTTAACTCCG 1731  
DB 1627 TCAGAGCTTCAGATTAATAACGCTGTTAATTTAAAGCTTGAAGCTACGATTAACACTCCG 1686  
QY 1732 TCAGAGCTTCGATACGATTTACGCTGATCATTCATCGGAGCGGATTTGCGCGAGCA 1790  
DB 1687 TCTAGATACCGTATCGGTTTCATGCACATTTGGATTCGANTGAATCGTTGATCA 1745

## RESULT 5

AAD09400

ID AAD09400 standard; cDNA; 1815 BP.

AC AAD09400;

XX 10-SEP-2001 (first entry)

XX Zea mays neoxanthin cleavage enzyme, VP14 cDNA.

XX Maize; neoxanthin cleavage enzyme; VP14; abscisic acid, ABA; herbicide;  
KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;  
KW plant growth protectant; ss.

XX Zea mays.

Key Location/Qualifiers  
CDS 1..1815  
FT /\*tag= a  
ET /product= "Zea mays VP14 protein"

XX EP1116794-A2.

XX 18-JUL-2001.

XX 11-JAN-2001; 2001EP-0300218.

XX 13-JAN-2000; 2000JP-0010056.

XX 11-JAN-2001; 2001JP-0003476.

XX (RIKE ) RIKEN KK.

XX Iuchi S, Kobayashi M, Shinozaki K;

XX WPI; 2001-400081/43.

XX P-PSDB; AAE04788.

XX A DNA encoding a protein with a neoxanthin cleavage activity for  
producing transgenic plants with improved or decreased stress tolerance

XX Claim 3; Page 60-64; 101pp; English.

XX The invention relates to neoxanthin cleavage enzymes and their  
corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key  
role in endogenous abscisic acid (ABA) biosynthesis under drought stress.  
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a  
CC plant when expressed in a plant cell. The invention also relates to  
CC methods for increasing or decreasing stress tolerance in a plant by  
CC introducing the DNA into the plant, and a transgenic plant into which a  
CC neoxanthin cleavage enzyme is introduced. The improvement of stress  
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin  
CC cleavage enzyme genes are useful for producing transgenic plants. An arid

CC land can be improved by growing transformant weed for several years and  
 CC then removing the weed by specifically lowering stress tolerance in the  
 CC weed by inducing an inducible promoter. The present cDNA sequence encodes  
 CC Zea mays neoxanthin cleavage enzyme, VP14 protein related to the  
 CC invention.

XX SQ Sequence 1815 BP; 266 A; 689 C; 604 G; 256 T; 0 other;  
 Query Match 33.7%; Score 606.6; DB 22; Length 1815;  
 Best Local Similarity 63.5%; Pred. No. 4.3e-186;  
 Matches 1003; Conservative 0; Mismatches 549; Indels 27; Gaps 4;  
 246 CACTAAACAGATGATTTGTCAGAGAGCGGCGGAGCGTTCGAGCGCGCGGAGG 305  
 Db 240 CAAGAGCAGCTCACTTGTTCAGCGCGCGCGGCGCGCTCGACCGTTCAGGA 299  
 Qy 306 TTTCCTTGT-----CAGCCACAGAGAGCTACACCCGCTTCCTAAACCGCTGATCTAG 359  
 Db 300 AGGTTTGTGGCCCAACGTCTCGAGCGCGCCACCGGCTGCCAGCAGCGCCGACCGGC 359  
 Qy 360 TGTTCAAGATCCCGGAAATTTGCTCGGCTGAATGAACAGCGCGTCCGGGTATCTTCC 419  
 Db 360 CGTCAGATCCCGCACTTCGCGCGCGCTCGGAGAGCGCGCCCTGCACGAGCTCCC 419  
 Qy 420 GTCGTCGGAACCTTCCCGATTCATCAAGAGAGTGTATGTGCGCAACGAGCTAACCC 479  
 Db 420 CGTCTCCGCGCGATCCCGCTTCAICGACGGGTCTACGCGCGCAACGCGCCACCC 479  
 Qy 480 ACTTCAGAGCGGTCAGATACCACTTTCGACGAGACGATGATTTACGCGCT 539  
 Db 480 CTGCTTACACCCCGTCGCGCGGCAACCTTTCGACGCGACGCGATGTGTGACGCGCT 539  
 Qy 540 ---CAAAATCGAACACGCTTACGCTAGCTAGCTTGGCGTTTACTCAGACTAACCGGTT 596  
 Db 540 GCGGATACGCAACGCGCGCGCGAGTCTTACGCTTCCGCTTACGAGACCGCGCGCT 599  
 Qy 597 TGTTCAGGAACGTAATTTGGTTCGACCGGTTTTTCCCAAGCCATTCGCTGAGCTTCAGG 656  
 Db 600 GCGCGAGGCGCGGATCGCGCGCGCTTTCGCCAAGCCATTCGCGAGCTGCGACGG 659  
 Qy 657 CCACACCGGATTCGCGGACATCATCTATTCTACGCGAGCTGCGCGGCTATGATCGA 716  
 Db 660 GCACTCCGCGATTCGCGCGCTCGCGCTTCTACGCGCGCGCGCTGCGGCTCGTGA 719  
 Qy 717 CCGGCGACACGGAACCGGCTGAGCTAACGCGGTTTGTCTATTTCATGTCGCGTATT 776  
 Db 720 CCCTTGGCGGACCGCGGTGCGCAACGCGCGCTCGCTTCTACTTCAACGCGCGCTGCT 779  
 Qy 777 GGTATGTGCGAGGATGATTTACCTTACCAAGTTGATCATCTCCCAATGGAGATTAA 836  
 Db 780 CGCATGTCCGAGGACGACCTCCCTTACCACTCCGCGTGGCGGACGCGGACCTCGA 839  
 Qy 837 AACCGTTGTCGTTGATTTGATGGCAATTAGAATCCCAATGATGCCACCCGAA 896  
 Db 840 GACCTGCGCGGTACCACTTCGACGGGAGCTCGGCTGCGCATGATCGCGCACCCAA 899  
 Qy 897 AGTCGACCGGAAATCCGTTGAACTTTCGTTTAACTAGCTACGAGCTGCTTCAAGCCTTA 956  
 Db 900 GCTGGACCGCGCACCGGAGCTCCACGCGTTCAGCTACGAGCTATCAAGAGCGGTA 959  
 Qy 957 CTTAAATATCTCCGATTTCTACCGGACGGAATTAATCACCGGAGCTCGAGATTCAGCT 1016  
 Db 960 CCTCAAGTCTTACTTTAGGCCCGACGACCAAGTCCGACGAGTGGAGATCCCGT 1019  
 Qy 1017 TGATCGCCCAACGATGATGACGATTCGCGATTACAGAGAACTTCGTCGTGATCTGA 1076  
 Db 1020 GGAGCGCCACGATGATCCACGACTTCGCGCATCACCGAGAACTTCGTCGTTGTGCCGA 1079  
 Qy 1077 CCAGCAAGTCTTTTCAGTTCGCGAGATGATCCCGGTGGGTCTCCGCTGTTTACGA 1136  
 Db 1080 CCACCAAGTGTGTTCAGCTCCAGGAGATGCTCCGCGCGGCTCCCGCTGTCGTA 1139  
 Qy 1137 CAAGAACAGGTGCGAAGATTTCGGGATTTTAGACAAATACGCCGCAAGATTTCATCGAACAT 1196

Db 1140 CAAGAGAGAGAGCTGCGGTTTCGCGTCTCCCAAGCAGCGCGGACGCTCGGAGAT 1199  
 Qy 1197 TAAGTGGATTGATGCTCCAGATTGCTTCTGCTTCCATCTCTGGAACGCTTGGGAAGACC 1256  
 Db 1200 GCGTGGGTGACGTCGCGGACTCTTCTGCTTCCACTGTGGAACGCTGGAAGGACGA 1259  
 Qy 1257 AGAACAGATGAAGTCGTCGTGATAGGTCTCTGATGACTCCACAGACTCAATTTTCAA 1316  
 Db 1260 GCGACGCGCGAGTGGTGTGATCGGCTCTCTGATGACCCCGCGGACTCATCTTCAA 1319  
 Qy 1317 CGAGTCTCAGAGAAATCTCAAGAGTGTCTCTGTAATTCGCTTGAATCTCAAAACCGG 1376  
 Db 1320 CGAGTCCGACGAGCGCTTGGAGAGCGTGTGACCGAGATCCGCTTGGACGCGCACGGG 1379  
 Qy 1377 TGAATCAACTCGCTCGATCATCTCCACAGAGATCAACAAGTCAACCTTCGAAGCAGG 1436  
 Db 1380 CCGGTTCACGCGCGCGCGCTCTGCTCC---GCCGTGCGAGGAGAACTTGGAGTGGG 1436  
 Qy 1437 GATGTCACAGAAACATGCTCGGCGGTAAACCAAAATTCGCTTACTTTGGCTTTAGCCGA 1496  
 Db 1437 CATGTGAACGCAACCTGCTGGCGCGGAGAGCGGTACGCTACCTCGCGGTGGCGGA 1496  
 Qy 1497 GCGTGGCTTAAAGTCTCAGGATTCGCTAAAGTTGATCTCACTACTGGAGAAAGTTAAAG 1556  
 Db 1497 GCGTGGCGCAAGGAGTCCGGCTTCGCCAAGAGAGACTGTCCAGCGCGAGCTCACCA 1556  
 Qy 1557 ACATCTTACGCGGATTAACCGTTACGAGGAGAGCTCTGTTTCTCCCGGAGAGAG- 1615  
 Db 1557 GTTCGAGTACGCGAGGCGCGTTCGCGCGGAGCGCTGCTTCTCCATCGACCCGCG 1616  
 Qy 1616 -----GAGAGGAAGAGAGGATACATCTCTACTTGGAGAAAGTTAAAG 1661  
 Db 1617 CCGCGGCCACCGCGCGGAGAGCGGTGACGTGCTTACCTTCGCTTCCATCGACCGCGG 1676  
 Qy 1662 GACATGGAATCGGAGTTACAGATAGTTAAACGCGCTTAGCTTAGAGTTGAAGCAACGCT 1721  
 Db 1677 CCGCGCACTCGGAGTACTTGTGTCAATCCGCGGACATCCGCTGGAGGCCACGCT 1736  
 Qy 1722 TAAATTCGTCAGAGGTTCCGTAAGGATTCAGGATTTACGCTTACGAGCGGAGCGATGATT 1781  
 Db 1737 TCAGTCCGCTCCCGGCTTCCGCTTCCAGGCACTTTCATCACGCGGCGAGGCT 1796  
 Qy 1782 GCGAAGCAGGTCGCTGA 1800  
 Db 1797 CGAGGCCAGGCGGCTGA 1815

## RESULT 6

AD09398  
 ID AD09398 standard; cDNA; 1734 BP.

XX AD09398;

AC AD09398;

DT 10-SEP-2001 (first entry)

XX Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED5 cDNA.

DE Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED5; abscisic acid; ABA; herbicide;  
 KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;  
 KW plant growth protectant; ss.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

PH Key Location/Qualifiers

FT CDS 1..1734

FT /\*tag= a

FT /product= "Arabidopsis thaliana AtNCED5 protein"

XX EP1116794-A2.

XX 18-JUL-2001.

PF 11-JAN-2001; 2001EP-0300218.  
XX  
PR 13-JAN-2000; 2000JP-0010056.  
PR 11-JAN-2001; 2001JP-0003476.  
XX  
XX (RIKE ) RIKEN KK.  
XX  
XX  
PI Tuchi S, Kobayashi M, Shinozaki K;  
XX WPI; 2001-400081/43.  
DR P-PSDB; AAE04786.  
XX  
XX  
PT A DNA encoding a protein with a neoxanthin cleavage activity for  
PT producing transgenic plants with improved or decreased stress tolerance  
XX  
XX  
XX Claim 3; Page 46-49; 101pp; English.  
XX  
XX The invention relates to neoxanthin cleavage enzymes and their  
CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key  
CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.  
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a  
CC plant when expressed in a plant cell. The invention also relates to  
CC methods for increasing or decreasing stress tolerance in a plant by  
CC introducing the DNA into the plant, and a transgenic plant into which a  
CC neoxanthin cleavage enzyme is introduced. The improvement of stress  
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin  
CC cleavage enzyme genes are useful for producing transgenic plants. An arid  
CC land can be improved by growing transformant weed for several years and  
CC then removing the weed by specifically lowering stress tolerance in the  
CC weed by inducing an inducible promoter. The present cDNA sequence encodes  
CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED5 protein.  
CC The AtNCED5 cDNA is obtained from an Arabidopsis plant-derived cDNA  
CC library using a cDNA of the CPRD65 (CowPea Responsive to Dehydration)  
CC gene isolated from cowpea plant as a probe.  
XX  
XX Sequence 1734 BP; 464 A; 390 C; 441 G; 439 T; 0 other;  
XX  
XX Query Match 29.3%; Score 528.2; DB 22; Length 1734;  
Best Local Similarity 61.0%; Pred. No. 1.4e-160;  
Matches 920; Conservative 0; Mismatches 563; Indels 24; Gaps 3;  
XX  
XX 306 TTTCTTGTGTCAGCCACGAGAGCTACACCGCTTCTCTAAACGGGTGATCTAGTTTCA 365  
DB 240 TATCGTTATTCCTATGGAGCAGAAATCGCCCGCTTCTTAAACGACCGCCGGGTTTCA 299  
XX 366 GATCGCGGAAATTTGCTCCGGTGAATGAACAGCCCGTCCGCGTAACTTCCCGTGT 425  
DB 300 ATTATCAGGTAACTTCGCTCCGGTTAATGAATGTCGGTTTCAACCGGTTTGAAGTGT 359  
XX 426 CGGAARACTTCCGATTCATCAAGAGGTATGTGCGCAACGAGCTAACCCACTTCA 485  
DB 360 TGGTCAGATTCCTTCTTGTCTAAAGGAGTTTACATCCGTAACGGTGCAACCTTAGTT 419  
XX 486 CGACCGGTTGACAGGTACCACTTCTTCGACGAGAGCGTATGTTTCAGCCGTCAAAT 545  
DB 420 TCCGCGGTAGCCGACATCATTTATTTGACGGTGACGGAATGATTCAGCCCGTTAGTAT 479  
XX 546 CGACACAGGT---TCAGTAGTACGTTGCGGTTTACTCAGACTAACCGGTTTCTTCA 602  
DB 480 CGGTTTGTAAACACAGGTAGTTACAGTGCCTGCTACACTAAACAAACCGGTTTGTCA 539  
XX 603 GGAACGTCATTTGGGTGACCGGTTTTCGCCCAAGCCATCGGTGAGCTTCACGCGCCACAC 662  
DB 540 AGAAACCGCGCTTGGACGATCGGTTTTCCTTAACCAATCGGCGAGCTTCACGCGCATTC 599  
XX 663 CGGTATTCGCCGACTCATGCTATTCTACGCCAGAGCTGCAGCCGGTATAGTACACCGGC 722  
DB 600 CGGTCTAGCTCGACTCGCTCTCTTCAACGGCTCGAGCTGGGATCGGTCTAGTGACCGGAC 659  
XX 723 ACACGGAACCGGTGTAGCTAACCGCGGTTGGTCTATTTCATGGCCGGTTATTGGCTAT 782  
DB 660 ACGTGGCATGGCGGTAGCTAACCGCGGTTGGTGTGTTTCTTTTAAACGGCAGGTTATTAGCCAT 719

QY 783 GTCCGAGGATGATTACCTTACCAAGTTCCAGATCACTCCCAATGGAGATTTAAACCGT 842  
DB 720 GTCCGAGGATGATCTTCTTACCAAGTTGAAGTTCGACGGTCAAGGAGATCTTTGAGACGAT 779  
XX 843 TGGTCGGTTCCGATTTTGGATGGCAATAGAAATCCCAATGATTCGCCACCCCGAAAGTCGA 902  
DB 780 CGACCGGTTCCGATTCGATGACCAAGATTCATCTTTCAGTGATAGCGCATCTTAAGTGGGA 839  
XX 903 CCAGGAAATCCGGTGAATCTCTTTCGCTTAAAGTTCAGACGCTGGTTTCAAAGCTTACCTAA 962  
DB 840 CGGACCCACAGGAGATCTCCATACACTGAGCTACAAACGTTTGAAGAAACCTCATCTCAG 899  
XX 963 ATACTTCGATTCCTCACCGGACGGAATAAATCACCGGACGCTCGAGATTCAGCTTCATCA 1022  
DB 900 GTATCTAAATTCACACGCTGCGGGAAGAACACGCTGACGTGGAGATCACGCTCCCTGA 959  
XX 1023 GCGAACGATGATGACAGATTTTCGGATTTACAGAAATCTTCGTCGTCGTCGTCGTCGTCGTC 1082  
DB 960 ACCAACGATGATTCATGATTTTCGGATAACCGGAAATTTTGTGTTTATACCGGATCAGCA 1019  
XX 1083 AGTCGTTTTCAGCTGCGCGGAGATGATCCGCGTGGTCTCTCCGGTGGTTCACGACAGAA 1142  
DB 1020 AATGGTATTCAAATTCGAAATGATTCGGGCGGCTCACCGGTTATCTACGTTAAAGA 1079  
XX 1143 CAAGGTCGCAAGATTCGGGATTTTAGACAAATACGCCGGAAGATTCATCGAACATTAAGTG 1202  
DB 1080 AAAAATGCGGAGATTTGGAGTTTGTCAAAGCAGGATCTGACCGGTCGGATATAAATTG 1139  
XX 1203 GATTGATGCTCCAGATTTGCTTCTTCATCTCTGAAACGCTTGGAAAGCCAGCAAAAC 1262  
DB 1140 GGTGATGATCCGATTTGTTTCTGTTTTCATCTATGGAATGCGTGGGAGAGAGACCGCA 1199  
XX 1263 AGATGAAG-----TCGTCGTGATPAGGTCCTGTATGACTCCACGAGACTCAATTTT 1313  
DB 1200 AGAGGAGAGACCGATTCGTCGTAATCGGTCATGATGAGCCACCCGACGATCTT 1259  
XX 1314 CAACGACTCTGACAGAAATCTCAAGAGTCTCTGTCGTAATCCGCTGATCTCAAAAC 1373  
DB 1260 TAGTGAATCAGGAGAACCAACCCGGGTTGAATTAAGTGAGATCCGGTTAAACATCGGTAC 1319  
XX 1374 CGGTGAATCAACTCCGCGTCCGATCATCTCCAAACGAAAGATCAACAAGTCAACCTCGAAGC 1433  
DB 1320 AAAAGATCGAAACCGTAAGGTTATCGTAAAC-----GGAGTGAATTTAGAACG 1367  
XX 1434 AGGATGGTCAACAGAAACATGCTCGCGCGTAAACCAATTCGCTTACTTGGCTTTAGC 1493  
DB 1368 GGGTCACATAAACCGTATGTTACGTTGGCCCGGAAAGCCAGTTCGTTTACATAGCAATAGC 1427  
XX 1494 CGACCGGTGGCTAAAGTCTCAGGATTCGCTAAAGTTGATCTCACTACTGGAGAAGTTAA 1553  
DB 1428 CGATCTTGGCCCAATTCAGTGGCATTCGAAAGTAGATATACAAACCGCACCGCTTTC 1487  
XX 1554 GAAACATCTTTACCGCGATTAACCGTTACGAGGAGAGCCCTCTGTTTCTCCCGGAGAGG 1613  
DB 1488 AGAGTTTAAATTAACCGACCGGCGGTTCCGGTGGAGAACCCGCTGTTTGTATCCGAGGGAGA 1547  
XX 1614 AGGAGAGGAAGACGAAGGATACATCTCTCTGTTTTCAGCGAGAGAGACATGGAAATC 1673  
DB 1548 AGGAGAGACACAAAGGTTATGTAATGGGTTTGTGAGAGACGAAGAGAACGAGTC 1607  
XX 1674 GGAGTTACAGATAGTTTAACCGCGTTAGCTTAGAGGTTGAAGCAACCGTTAACTTCGTC 1733  
DB 1608 GGAGTTTGTGGTGTGCGACGCGAGGATATGAAGCAAGTCCGCGCGTGGCTTTCGCGGA 1667  
XX 1734 AAGGTTCCGTTACGATTTTACCGGTACATTCATCCGAGCCGATGATGTTTGGCGAAGCAGGT 1793  
DB 1668 GAGGATCTTATGGTTTCCATGAAACGTTCTGAGCGAGATCAGTTGAAGCAACAGT 1727  
XX 1794 CGTGGA 1800  
DB 1728 TTTCTGA 1734

```

RESULT 7
AAC56678
ID AAC56678 standard; DNA; 443 BP.
XX
AC AAC56678;
XX
DT 25-JAN-2001 (first entry)
XX
DE Eucalyptus grandis transcription factor DNA sequence #549.
XX
KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW homeodomain zipper; LIM domain; AP2; ERBBs; zinc finger domain;
KW type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS Eucalyptus grandis.
XX
PN WO200053724-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US06112.
XX
PR 11-MAR-1999; 99US-0266513.
XX
PR 18-AUG-1999; 99US-0149485.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Wood M, McGrath A, Shenk MA, Glenn M;
XX
WPI; 2000-579369/54.
XX
New isolated polynucleotide encoding a plant transcription factor for
producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide
PT
XX
Claim 1; Page 490; 747pp; English.
XX
The present invention relates to novel plant transcription factors from
Eucalyptus grandis or Pinus radiata. The present sequence is the coding
sequence for one such transcription factor. The transcription factor may
be used to produce a plant having modified gene expression such as a
woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
mahogany species or to modify the activity of a polypeptide in a plant.
The transcription factors of the present invention are members from the
following families of regulatory proteins: bZIP, bZIP family of G-box
binding factors, basic helix-loop-helix zipper, LIM domain, AP2
homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
and ERBBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
and MYB.
XX
Sequence 443 BP; 71 A; 177 C; 117 G; 77 T; 1 other;
SQ
Query Match 12.8%; Score 229.8; DB 21; Length 443;
Best Local Similarity 69.9%; Pred. No. 7.9e-64;
Matches 309; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
XX
620 GACCGGTTTCCCAAGCCATCGGTGAGCTTCACGGCCACACCGGTATTGCCGACTCA 679
1 GCCCGCTTCNCNAGGCCATCGGAGCTCCACGGCCACTCCGGCATCGGCGCTCA 60
680 TGTATTCTACGCCAGCTGAGCCGGTATAGTCAGCCGCCACACCGGATGGCGTGTAG 739
61 TGTCTTCTACGCCCGCAGCCCTTTCGGCTCTGTCGACCAACCGGAATGGCATGGCGCTCG 120
740 CTAAACCGGTTTGGTCTATTTCATATGCGCGGTATTGCTATGTCGAGGATGATTAC 799
121 CGAACCGCGGCTCTGTGTACTTCGACGGCCACTCTCTCGCGATGTCGAGACGACCTCC 180

```

```

800 CTTACCAAGTTTCAATCACTCCCAATGGAGATTAAAAACCGTTGGTGGTTGATTTTG 859
181 CCTACCACGTCGCGTCACGCGCTCCGGCGACTCGAGACCGTCGCGCGCTAGCACTTCG 240
860 ATGGACAAATTAGAAATCCAAATGATTCGCCACCCGAAAGTCGACCCGGAATCCGGTGAAC 919
241 CCGGCCAGCTCGACTCTCCGATGATCGCCACCCGAGATCGACCGGTTCCGGCGAGA 300
920 TCTTCGCTTTAAGCTAGACGTCGTTTCAAGCTTAAAGCTTACCTAAATACTTCCGATTTCCAC 979
301 TGTTCGCCCTCAGCTACGACGTCGTCGGAAGCGGTACCTCAAGTACTTCCGATTTCCCA 360
980 CGGACCGAACTAATCACCAGGAGCTGAGATTTCAGCTTGCATCAGCCACGATGATGCACG 1039
361 AGGACGGCGAGAAAGTCTCCCGAGCTGAGATCCCTGCTGCTGAGCGACCAATGATGCAG 420
1040 ATTTCGCGATTACAGAACTT 1061
421 ATTTCGCCATCACCAGCGCTT 442
RESULT 8
AAC82706/c
ID AAC82706 standard; DNA; 492 BP.
XX
AC AAC82706;
XX
DT 15-MAR-2001 (first entry)
XX
DE Rice abscisic acid synthesis associated DNA SEQ ID NO 5.
XX
KW Rice; abscisic acid synthesis; VPL4-like gene; drought resistance; ds.
XX
OS Oryza sativa.
XX
PN WO200071727-A1.
XX
PD 30-NOV-2000.
XX
PF 25-MAY-1999; 99WO-JP02734.
XX
PR 25-MAY-1999; 99WO-JP02734.
XX
PA (NORQ) JAPAN MIN AGRIC FORESTRY & FISHERIES.
PA (BIOO-) BIO-ORIENTED TECHNOLOGY RES ADVANCEMENT.
XX
Hirochika H, Sakamoto K;
WPI; 2001-032042/04.
XX
Oligonucleotide encoding gene for regulating abscisic acid synthesis in
plants, useful for constructing e.g. genetically-modified rice with
drought resistance and ear-germination resistance -
XX
Disclosure; Page 50; 55pp; Japanese.
XX
This invention describes a novel polynucleotide sequence (I) which
encodes a protein capable of regulating the synthesis of abscisic acid.
The invention also describes (1) an oligonucleotide encoding a protein
of a gene obtained by controlling the expression of a VPL4-like gene;
(2) a vector containing the oligonucleotide ligated operably to the
regulation sequence; (3) a plant transformed with the vector; and
(4) a method for regulating abscisic acid synthesis in a plant including
the transfer of the above oligonucleotide to it. The gene is useful for
constructing drought resistant rice.
XX
Sequence 492 BP; 90 A; 126 C; 186 G; 88 T; 2 other;
SQ
Query Match 11.3%; Score 204; DB 22; Length 492;
Best Local Similarity 69.7%; Pred. No. 2.1e-55;
Matches 304; Conservative 0; Mismatches 130; Indels 2; Gaps 2;
XX
627 TTTCCTCCCAAGCCATCGGTGAGCTTCACGGCCACACG-GTATTGCCGACTCATCTAT 685

```





PN WO200053724-A2.  
 XX 14-SEP-2000.  
 XX 09-MAR-2000; 2000WO-US06112.  
 XX 11-MAR-1999; 99US-0266513.  
 PR 18-AUG-1999; 99US-0149485.  
 XX (GENE-) GENESIS RES & DEV CORP LTD.  
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
 XX Wood M, McGrath A, Shenk MA, Glenn M;  
 XX WPI; 2000-579369/54.  
 XX New isolated polynucleotide encoding a plant transcription factor for  
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
 PT having modified gene expression or modified activity of a polypeptide  
 PT -  
 XX Claim 1; Page 494; 747pp; English.  
 XX The present invention relates to novel plant transcription factors from  
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding  
 CC sequence for one such transcription factor. The transcription factor may  
 CC be used to produce a plant having modified gene expression such as a  
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or  
 CC mahogany species or to modify the activity of a polypeptide in a plant.  
 CC The transcription factors of the present invention are members from the  
 CC following families of regulatory proteins: bZIP, bZIP family of G-box  
 CC binding factors, basic helix-loop-helix zipper,  
 CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2  
 CC and ERBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements  
 CC and MYB.  
 XX Sequence 372 BP; 60 A; 148 C; 101 G; 63 T; 0 other;  
 SQ  
 Query Match 10.4%; Score 188; DB 21; Length 372;  
 Best Local Similarity 69.1%; Pred. No. 2.9e-50;  
 Matches 257; Conservative 0; Mismatches 115; Indels 0; Gaps 0;  
 QY 639 CATCGTGAGCTTCAGGCCACACCGGATATGCCGACTCATGCTATTTCTACGCGAGAGC 698  
 Db 1 CATCGGAGCTTCCAGGCCACTCCGGCATCGCGGCTCATGCTCTTTCTACGCGCGAG 60  
 QY 699 TGCAGCCGGTATAGTCGACCCCGGCACACGGAACCGGTGTAGCTAACGCGGTTTGGTCTA 758  
 Db 61 CTTCTTCGGCTCTGTCGACCCCGGATGGCATGGCTCGGAGCGCGCTCGTCTA 120  
 QY 759 TTTCAATGCGCGGTTATTTGGCTATGTCGGAGATGATTTACCTTACCAAGTTTCAGATCAC 818  
 Db 121 CTTGACGGCCACCTCCTCGCGATGTCGGAGGACGACCTCCCTACCACTGCGCGTCC 180  
 QY 819 TCCCAATGGAGATTTAAAGACCGTTGGTTCGATTTTGTATGGACAAATTAGATCCAC 878  
 Db 181 GGGCTCCGGCGACTTCGAGACCGTTCGGCGCTACGATTCGCGCGGAGCTCGACTTCC 240  
 QY 879 AATGATGCCCCACCCGAAAGTCAGACCCGGAATCCGGTGAATCTTTTCGTTTAACTACGA 938  
 Db 241 GATGATCGCCACCCGGAAGATCGACCCGGCTTCGCGGAGATGTTTCGCCCTCAGCTACGA 300  
 QY 939 CTTCTTTCAAGCCCTTACCTAAATACCTTCGATTTCTCAGCGAGCAATTAATCAC 998  
 Db 301 CTTCTTCGGAAGCGGTACCTCAAGTACTTCCGATTTCTCAAGGAGCGCGAGAGTCCCC 360  
 QY 999 GGACGTCGAGAT 1010  
 Db 361 CGACGTCGAGAT 372

RESULT 11  
 AAC56548

ID AAC56548 standard; DNA; 325 BP.  
 XX AAC56548;  
 XX 25-JAN-2001 (first entry)  
 DE Eucalyptus grandis transcription factor DNA sequence #419.  
 XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;  
 KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;  
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;  
 KW homeodomain zipper; LIM domain; AP2; ERBs; zinc finger domain;  
 KW type 2 Cys2His2; CCAAT box element; MYB; ss.  
 XX Eucalyptus grandis.  
 CS WO200053724-A2.  
 XX 14-SEP-2000.  
 XX 09-MAR-2000; 2000WO-US06112.  
 XX 11-MAR-1999; 99US-0266513.  
 PR 18-AUG-1999; 99US-0149485.  
 XX (GENE-) GENESIS RES & DEV CORP LTD.  
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
 XX Wood M, McGrath A, Shenk MA, Glenn M;  
 XX WPI; 2000-579369/54.  
 XX New isolated polynucleotide encoding a plant transcription factor for  
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
 PT having modified gene expression or modified activity of a polypeptide  
 PT -  
 XX Claim 1; Page 461; 747pp; English.  
 XX The present invention relates to novel plant transcription factors from  
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding  
 CC sequence for one such transcription factor. The transcription factor may  
 CC be used to produce a plant having modified gene expression such as a  
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or  
 CC mahogany species or to modify the activity of a polypeptide in a plant.  
 CC The transcription factors of the present invention are members from the  
 CC following families of regulatory proteins: bZIP, bZIP family of G-box  
 CC binding factors, basic helix-loop-helix zipper,  
 CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2  
 CC and ERBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements  
 CC and MYB.  
 XX Sequence 325 BP; 45 A; 128 C; 100 G; 52 T; 0 other;  
 SQ  
 Query Match 9.5%; Score 171.8; DB 21; Length 325;  
 Best Local Similarity 71.2%; Pred. No. 5e-45;  
 Matches 227; Conservative 0; Mismatches 92; Indels 0; Gaps 0;  
 QY 451 GGAGTGATGTGCGCAACCGAGCTAACCCATTCACGAGCGCGGTGACAGGTCAACACTTC 510  
 Db 3 GGGGTGTAGCTTCGGCAACCGCGCCACACCGCTCCACGAGCGGTCCCGGGCACCACTTG 62  
 QY 511 TTCGCGGAGCGGTATGTTTCACGCGGCAAAATTCGAACACGCGTTTCAGCTAGCTACGCT 570  
 Db 63 TTCGCGGAGCGGATGATTCACGCGGTTCGGGTTTCGGGGGCTTCAGTGAGCTACGCC 122  
 QY 571 TGCCGGTTTACTCAGACTAACCGGTTTGTTCAGGAACGTCATTTGGGTTCGACCGGTTTC 630  
 Db 123 TGCCGGTTTACCGAGACGCAACCGCTGATCCAGGAACCGGGCTCGCGCGCACCACTTG 182  
 QY 631 CCCAAGCCCATCGGTAGCTTACGCCACACCGGTATTCGCCGACTCATGTATTTCTAC 690  
 Db 183 CCCAAGCCCATCGCGGATCCACGCCCACTCCGCGATCCGCGGCTCATGCTCTTCTAC 242

QY 691 GCGAGGTGCGCGGTATAGTCACCGCGGCACACGGAACCGGTGTAGCTAACCGGT 750  
Db 243 GCGCGCGCTCTTCGCGCTCGTCGACACCGTAATGCGATGCGGTGCGGAACGCGGC 302  
QY 751 TTGGTCTATTTCAATGGCC 769  
Db 303 CTCGTGACTTCGACGCC 321

RESULT 12  
AAC42989  
ID AAC42989 standard; DNA; 1788 BP.  
XX AC AAC42989;  
XX XX  
DT 17-OCT-2000 (first entry)  
XX XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 37611.  
XX XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX XX  
OS Arabidopsis thaliana.  
XX XX  
PN EP1033405-A2.  
XX XX  
PD 06-SEP-2000.  
XX XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127452.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136382.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138647.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 02-AUG-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.

PR	10-AUG-1999;	99US-01481171.	Db	393	CGGCACTCTTCCACTGTCACTTAACGGCGCTTATCATCGTAAACGGTCCAAATCCACAGTT	452
PR	11-AUG-1999;	99US-01483119.	Qy	486	CGAGCCGGTGACAGGTCAACCACTTCTTCGACGGAGACGGTATGTTTCACGGCGTCAAAAT	545
PR	12-AUG-1999;	99US-01483411.	Db	453	TCTCCCTCGTGGTCTTACCATCTCTTCGACGGGACGGTATGCTTTCACGGCATAAAT	512
PR	13-AUG-1999;	99US-0148684.	Qy	546	CGAACACGGTTCACTAGTACGCTTGGCGGTTTACTCAGACTAACCGGTTGTTTCAGA	605
PR	16-AUG-1999;	99US-01491175.	Db	513	CCACAACGGTAAAGCACTCTCTGTAGCAGATACGTCAGACTTATAAATCAACGTCGA	572
PR	17-AUG-1999;	99US-0149426.	Qy	606	ACGTCAATTTGGGTGACACGGGTTTCCCAAGGCCATCGGTGAGCTTTCACGGCCACACCG-	664
PR	20-AUG-1999;	99US-0149723.	Db	573	GAACAAACCGGAGCTCGGTTATGCTTAACTGCTTTTCCGGATTCAACGGTGTAAACGC	632
PR	23-AUG-1999;	99US-0149902.	Qy	665	--GTATTGCCCACTCATGCTATTCTACGCCAGACTGACGGGTATAGTCGACCCGGC	722
PR	25-AUG-1999;	99US-0150566.	Db	633	GTCACTAGTCTCGTAGGCTTTTAAACGGCAGCTAGGCTTTTAAACGGCAGTATAATCCCGG	692
PR	26-AUG-1999;	99US-0150884.	Qy	723	ACAGGGAACCGGTAGCTAACCGGTTTGGTCTTATTTCAATGGCCGGTTATTTGGCTAT	782
PR	27-AUG-1999;	99US-0151066.	Db	693	TAACGGCATGTTTGTAGCTAATCAAGCTAGCTTTCTTCAGTAACCGCTCTCTTTGCTTT	752
PR	27-AUG-1999;	99US-0151080.	Qy	783	GTCCGAGGATGATTACCTTACCAAGTTTCAGATCATCTCCCAATGAGATTTAAACCGT	842
PR	30-AUG-1999;	99US-0151309.	Db	753	AGTGAATCTGATTTTACCTTACCGGTCGATTAACCGAATCAGGAGATATTGAACCGAT	812
PR	31-AUG-1999;	99US-0151338.	Qy	843	TGGTGGTTTGGATTTTGTAGTGAACAATTAGATTCACCAATGATTGCCACCCGAAAGTCGA	902
PR	01-SEP-1999;	99US-0151930.	Db	813	CGACCGGTACGATTTTCGACGGGAATTAACGGATGATGATGACAGCTCATCTTAAACCGA	872
PR	07-SEP-1999;	99US-0152363.	Qy	903	CCCGAATCCGGTGAATCTTCCGTTTAAAGTTCAGACGTCGTTTCAAGCCCTTACCTAAA	962
PR	10-SEP-1999;	99US-0153070.	Db	873	TCCAAATAACCGGAGAACTTTCGCTTCCGGTACGGTCCGGTTCCA---CCGTTTTTAAC	929
PR	13-SEP-1999;	99US-0153758.	Qy	963	ATACTTCCGATCTCACCGGACGGAACTAAATCACCGGACGTCGAGA---TTTCAGTTGA	1019
PR	15-SEP-1999;	99US-0154018.	Db	930	ATATTTCGGTTGATTCGCGCGGGAATAAACAAGAGACGTTCCGATTTCTCGATGAC	989
PR	16-SEP-1999;	99US-0154039.	Qy	1020	TCAGCCAAAGATGATGACACGATTTCCGATTTCAGAGAACTTCGTCGTCTACCTGACCA	1079
PR	20-SEP-1999;	99US-0154779.	Db	990	GTCTCCGTCTTCTCCATGACTTCGCGATCACGAAACGTCACCGGATTTTCGACAGAT	1049
PR	22-SEP-1999;	99US-0155139.	Qy	1080	GAAAGTCG-----TTTTCAAGTCGCGGAGATGATCCGCGGTGGTCTCCGGTGT	1130
PR	23-SEP-1999;	99US-0155486.	Db	1050	TCAGCTTGGCATGAGGATGAACATGTTGGATTGTTCTCGAAGTGGTCTTCCGGTTGG	1109
PR	24-SEP-1999;	99US-0155659.	Qy	1131	TTACGACAAGAACAGGTCCGAAGTTCGGGATTTTAGCAAAATACGCCAGAGATTTCATC	1190
PR	28-SEP-1999;	99US-0156458.	Db	1110	TACTGATACGGAAACTCCAGGCTTGGAGTGATTCCTAAGTACGCCGAGATGAGTC	1169
PR	29-SEP-1999;	99US-0156596.	Qy	1191	GAACATTAAAGTATGATGCTCCAGATGCTTCTGCTTCCTCCATCTCTGGAACGTTGGGA	1250
PR	04-OCT-1999;	99US-0157117.	Db	1170	GGAGATGAATGGTTCGAAATTCCTCGGATTCATATCATTCAGCTATTAAATGCTTGGGA	1229
PR	05-OCT-1999;	99US-0157753.	Qy	1251	AGAGCCAGAAACAGATGAAGTCTCGTGATAGGCTCTGATGACTCCACAGACTCAAT	1310
PR	06-OCT-1999;	99US-0157865.	Db	1230	TGAAGATGATGAAACAGCGTCTTTTGGATTGACCCGAAATATTATGTCGATTGAACATC	1289
PR	07-OCT-1999;	99US-0158029.	Qy	1311	TTTCAACGAGTCTGACGAGAACTCAAGAGTGTCTCTGTCTGAAATCCGCTGATCTCAA	1370
PR	12-OCT-1999;	99US-0158232.	Db	1290	TTTAGAGAGGATGATCTGGTTC--ATGCTTTGGTGGAGAGGTGAAGATCGATCTCGT	1346
PR	13-OCT-1999;	99US-0158293.	Qy	1371	AACCGGTGAATCAACTCGCGCTCCGATCATCTCCAAAGAGATCAACAGTCAACCTCGA	1430
PR	13-OCT-1999;	99US-0158295.	Db	1347	CACCGGATTTGAGAGCTCATCCGATCTCAGGAGG-----AATCTCGA	1391
PR	14-OCT-1999;	99US-0159331.	Qy	1431	AGCAGGATGTTCAACAGAAACATGCTCGCCGCTAAACAAATTCGCTTACTTGGCTTT	1490
PR	14-OCT-1999;	99US-0159337.	Db	1392	TTTCGCTGTGATTAATCCGCGCTTCTCCGGAGATGTAGCAGTACGTTTACGCGCGAT	1451
PR	18-OCT-1999;	99US-0159637.	Qy	1491	AGCCGACCGTGGCTTAAAGTTCAGATTCGCTTAAAGTTGATCTCACTACTCTGAGAAGT	1550
PR	21-OCT-1999;	99US-0160741.	Db	1452	TGGAGATCCGATGCCGAAGATCTCCGGTGTGGTGAAGCTTGTGTCTTAAAGAGATCG	1511
PR	21-OCT-1999;	99US-0160767.				
PR	21-OCT-1999;	99US-0160768.				
PR	21-OCT-1999;	99US-0160770.				
PR	21-OCT-1999;	99US-0160814.				
PR	21-OCT-1999;	99US-0160815.				
PR	22-OCT-1999;	99US-0160880.				
PR	22-OCT-1999;	99US-0160981.				
PR	22-OCT-1999;	99US-0160989.				
PR	25-OCT-1999;	99US-0161404.				
PR	25-OCT-1999;	99US-0161405.				
PR	25-OCT-1999;	99US-0161406.				
PR	26-OCT-1999;	99US-0161359.				
PR	26-OCT-1999;	99US-0161360.				
PR	26-OCT-1999;	99US-0161361.				
PR	28-OCT-1999;	99US-0161920.				
PR	28-OCT-1999;	99US-0161992.				
PR	28-OCT-1999;	99US-0161993.				
PR	29-OCT-1999;	99US-0162142.				

Query Match 9.4%; Score 169.6; DB 21; Length 1788;  
Best Local Similarity 50.2%; Pred. No. 7.1e-44;  
Matches 709; Conservative 0; Mismatches 634; Indels 69; Gaps 9;  
426 CGGAAACTTCCCGATTCCATCAAGAGGTGATGTGCGCAACGGAGCTTAACCCCACTTCA 485

QY 1551 TAAGAAACATC-----TTTACGGGATAAACCGTTACGGAGGAGCCCTCT 1595  
 Db 1512 GATGATTGTACGGTGGCCGCTAGATGTACGGTTCAGGTTGTTACGGCGGAGAACCGTT 1571  
 QY 1596 GTTCTCCCGGAGAGAGGA-----GAGGAAGAGAGATACATCCCTG 1643  
 Db 1572 TTTCGTAGCTAGGGATCTCTGTTAATCCGAGCGGAGGAGGATGATGTTATGTTGAGTAC 1631  
 QY 1644 TTTCGTTCACGACGAGAGACATGAAATCGAGTTACAGATAGTTAAACGCCGTTAGCTT 1703  
 Db 1632 GTATGTTACAGATGAGTACGAGATCGAAGTTTCTGTTGATGGACGCTAAATGCC 1691  
 QY 1704 AGAGGTTGAA-----GCAACGGTTAACTTCGTCAGAGGTTCCGTACGAGTTTCACGG 1757  
 Db 1692 GGAGCTTGAATTCGTCCCGCGCTGAGGTTGCCGGAAGGGTTCGTACGGAATCCATGG 1751  
 QY 1758 TACATTTCATCGGAGCGGATGATTTCGCGAAGC 1789  
 Db 1752 GTTATTTGTCAGGAAGTGACCTTAATTAAGC 1783

## RESULT 13

AAD09395  
 ID AAD09395 standard; cDNA; 1788 BP.  
 XX  
 AC AAD09395;  
 XX  
 DT 10-SEP-2001 (first entry)  
 XX  
 DE Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED2 cDNA.  
 XX  
 KW Neoxanthin cleavage enzyme; AtNCED2; abscisic acid; ABA; herbicide;  
 KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;  
 KW plant growth protectant; ss.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PH Key Location/Qualifiers  
 FT CDS 1..1788  
 FT /\*tag= a  
 FT /product= "Arabidopsis thaliana AtNCED2 protein"  
 XX  
 XP EPI116794-A2.  
 XX  
 XX 18-JUL-2001.  
 XX  
 XX 11-JAN-2001; 2001EP-0300218.  
 XX  
 XX 13-JAN-2000; 2000JP-0010056.  
 XX  
 XX 11-JAN-2001; 2001JP-0003476.  
 XX  
 XX (RIKE ) RIKEN KK.  
 XX  
 XX Iuchi S, Kobayashi M, Shinozaki K;  
 XX  
 XX WPI; 2001-400081/43.  
 XX  
 XX P-PSDB; AAE04783.

A DNA encoding a protein with a neoxanthin cleavage activity for producing transgenic plants with improved or decreased stress tolerance

Example 10; Page 25-29; 101pp; English.

The invention relates to neoxanthin cleavage enzymes and their corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key role in endogenous abscisic acid (ABA) biosynthesis under drought stress. Neoxanthin cleavage enzyme is used for improving stress tolerance in a plant when expressed in a plant cell. The invention also relates to methods for increasing or decreasing stress tolerance in a plant by introducing the DNA into the plant, and a transgenic plant into which a neoxanthin cleavage enzyme is introduced. The improvement of stress

CC tolerance in plants is useful, for example in plant breeding. Neoxanthin cleavage enzyme genes are useful for producing transgenic plants. An arid land can be improved by growing transformant weed for several years and then removing the weed by specifically lowering stress tolerance in the weed by inducing an inducible promoter. The present cDNA sequence encodes Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED2 protein. The AtNCED2 cDNA is obtained from an Arabidopsis plant-derived cDNA library using a cDNA of the CPRD65 (Cowpea Responsive to Dehydration) gene isolated from cowpea plant as a probe.

XX Sequence 1788 BP; 444 A; 444 C; 418 G; 482 T; 0 other;

Query Match 9.4%; Score 169.6; DB 22; Length 1788;  
 Best Local Similarity 50.2%; Pred. No. 7.1e-44;  
 Matches 709; Conservative 0; Mismatches 634; Indels 69; Gaps 9;  
 QY 426 CGAAATCTCCGATCCATCAAGGAGTGTATGCGCAACGGAGCTAACCCACTTCA 485  
 Db 393 CGGCACTCTTCCACTGTCACTTAACCGCGCTTACATCCGTACGGTCCAAATCCACAGTT 452  
 QY 486 CGAGCCGGTGACAGGTCAACACTTCTTCGACGGAGACGGTATGCTTACGCCATAAAAT 512  
 Db 453 TCTCCCTCGTGGTCCCTTACCATCTCTTCGAGCGGACGGTATGCTTACGCCATAAAAT 512  
 QY 546 CGAACACGGTTCACTAGTACGCTTCCCGGTTTACTCAGACTTAACCGGTTTGTTCAGGA 605  
 Db 513 CCACACGGTAAAGCCACTCTCTGTAGCAGATACGTCACAGACTTATAATACACGTCGA 572  
 QY 606 ACGTCAATTGGGTGACCGCGGTTTTCCCAAAGCCATCGGTGAGCTTCACGGCCACACCG- 664  
 Db 573 GAACAAACCGGAGCTCGGTTATGCTTAACGGTGTTCGGATTCAACGGTGAACGGC 632  
 QY 665 --GTATTGCCCGACTCATGCTATTCTACGCCAGAGCTGACGGGTATAGTCGACCGCGC 722  
 Db 633 GTCAGTAGCTCGTGGAGCTTTAAACGGCAGCTAGGGTTTAAACCGGACAGTATTAATCCGGT 692  
 QY 723 ACACGGAACCGGTGTAGCTAAACCGCGTGTGGTCTATTTCATATGGCGGTTATTGGCTAT 782  
 Db 693 TAACGGCATTTGGTTTGTAGTAAATACAGTCTAGCTTCTTCTTCAAGTAAACCGTCTCTTTGCTTT 752  
 QY 783 GTCGAGGATGATTACCTTTACCAAGTTTACAGTACATCCCAATGGAGATTTAAACACCGT 842  
 Db 753 AGGTGAATCTGATTTACCTTACCGCTACCGCTCCGATTAAACCGAATACGAGATATTCGAACGAT 812  
 QY 843 TGGTCGGTTGATTTTGTATGAGCAATAGAAATCCAAATGATTCGCCACCGCAAGTCGA 902  
 Db 813 CGGACGGTACGATTTCCGCGGGAATTTAGGATGAGTATGACAGCTCATCTTAACACCGA 872  
 QY 903 CCGGAATCCGGTGAATCTTTCGCTTTAAGTACGACGTGCTTTTCAAAGCCTTACCTAAA 962  
 Db 873 TCCAATAACCGGAGAACTTTCGCTTTCCGGTACGGTCCGGTTCCA---CGGTTTTTAAC 929  
 QY 963 ATACTTCGGATTCACCGGACGGAACCTAAATCACCAGGAGCTCGAGA---TTCAGCTTGA 1019  
 Db 930 ATATTTCGGGTTGATTTCCGCCGGGAATTAACAAGAGAGCTTCGATATTTCTCGATGAC 989  
 QY 1020 TCAGCCAAACGATGATGACACGATTTCCGATTTACAGAGAACTTCGTCGTCGTAACCTGACCA 1079  
 Db 990 GTCTCCGTCGTTTCTCCATGACTTTCGGATTCACGAAACGTCACCGGATTTTCGAGAGAT 1049  
 QY 1080 GCAAGTCG-----TTTTCAAGCTCCCGAGATGATCCGCGTGGTCTCCGGTGGT 1130  
 Db 1050 TCAGCTTGGCATGAGGATGAACATGTTGGATTCTCGAAGGTGGTCTCCGGTTGG 1109  
 QY 1131 TTACACCAAGAAACAAAGGTCCGAAGATTCGGGATTTTAGACAAATACGCCGAAGATTTCATC 1190  
 Db 1110 TACTGATTAACGGMAAAACTCCAAAGCTTGGAGTGATTCCTAAGTACGCCGGAGATGAGTC 1169  
 QY 1191 GAACATTAAGTGGATGATGCTCCAGATGCTTCTGTTTCCATCTCTGGAACGCTTGGGA 1250  
 Db 1170 GGAGATGAATGGTTTCGAAGTTCCTGGATTCAATATCATTCACGGTATTAATGCTTGGGA 1229  
 QY 1251 AGAGCCAGAAACAGATGAAGTTCGTCGATAGGGTCTCTGTATGACTCCACAGACTCAAT 1310



QY 1131 TTACGACAAAGCAAGTGGCGAGATTTTGGGATTTTAGACAAATAGCGGAGATTCATC 1190  
 Db 1110 TACTGATACCGGAAACCTCAGGCTTGGAGTATCTTAAGTAGCGCGGAGTCACTC 1169  
 QY 1191 GAAATTAAGTGGATGATGCTCAGATTTGCTTCTTCCATCTCTGGAACGCTTGGGA 1250  
 Db 1170 GGAGATGAATGGTTCGAAGTCTCTGGATTCATATCAITTCACGCTATTAATGCTTGGGA 1229  
 QY 1251 AGAGCCAGAAACAGATGAAGTCGTCGTGATAGGTCCTGTATGACTCCACGACTCAAT 1310  
 Db 1230 TGAAGATGATGAACACGCTGCTTGTATGACCGAATATTATGTCGATTGAACATAC 1289  
 QY 1311 TTTCAACAGTCTCAGAGATCTCAAGAGTGTCTGTCTGAATCGGCTGATCTCAA 1370  
 Db 1290 TTTAGAGAGTGTATCTGGTTC---ATGCTTTGGTGGAGAGGTGAAGATCATCTGT 1346  
 QY 1371 AACCGGTGAATCAACTCGCTCGATCATCTCCACAGAGATCAACAAGTCAACCTCGA 1430  
 Db 1347 CACCGGATTTGAGAGCTCATCCGATCTCAGCGAGG-----AATCTCGA 1391  
 QY 1431 AGCAGGATGCTCAACAGAAACATGCTCGGCGGTAAACCAAAATTCGCTTACTTGGCTTT 1490  
 Db 1392 TTTCTGCTGTATTAATCCGGGTTTCTCGGAGATGTAGCAGTACGTTTACCGCGCAT 1451  
 QY 1491 AGCCGAGCGGTGGCTAAAGTCTCAGGATTCGCTAAAGTTGATCTCACTACTGGAGAAT 1550  
 Db 1452 TGGAGATCCGATGCCGAGATCTCCGCTGTGTGAGCTTGTATGCTTAAGAGGATCG 1511  
 QY 1551 TAAGAAATC-----TTACGCGGATAACCGTTAGGAGGAGGCTCT 1595  
 Db 1512 GATGATTTAGGTGGCCCTAGATGTACGGTTACGGTTGTTACGGCGGAGAACCGTT 1571  
 QY 1596 GTTCTCCCGGAGAGAGGA-----GAGGAAGACGAGGATACATCTCTG 1643  
 Db 1572 TTTCTAGTAGGATCTCGTAAATCCGAGGCGGAGGATGATGTTATGTGTGAC 1631  
 QY 1644 TTTCTTACGACGAGAGATCGAATCGGATTCAGATAGTTAAACGCTTAGCTT 1703  
 Db 1632 GTATGTTCCAGTGAAGTGAATCGAGATCGAAGTTCTGTTGATGGAGCTTAATTCGCG 1691  
 QY 1704 AGAGGTTGAA-----GCAACGTTTAACTTCGCTCAAGGTTCCGTAGCGATTTCA 1757  
 Db 1692 GGAGCTTGAATCGTCGCGCGCTGAGGTTGCCGCGAAGGTTCCGTAGCGATTTCCATCG 1751  
 QY 1758 TACATTCATCGGCGCGATGATTTGGCGAGC 1789  
 Db 1752 GTATTGTCAGGAAGTGAACCTTAATAGC 1783

## RESULT 15

AAF77206  
 ID AAF77206 standard; cDNA, 1950 BP.

AC AAF77206;

XX AAF77206;

AC AAF77206;

XX cDNA encoding sunflower neoxanthin cleavage enzyme (NCE).

XX Defence-related signalling gene; sunflower; neoxanthin cleavage enzyme;

XX NCE; amino acid permease; AAP; glutamic acid rich protein; GRP;

XX pathogen resistance; abscisic acid metabolism; ss.

XX Helianthus annuus.

OS Location/Qualifiers

XX Key 1..1632

XX CDS /tag= a

XX /product= "NCE"

XX /note= "Neoxanthin cleavage enzyme"

XX /partial

XX WO200112801-A2.

XX 22-FEB-2001.  
 PD 17-AUG-2000; 2000WO-US22961.  
 XX 18-AUG-1999; 99US-0149656.  
 PR 23-MAY-2000; 2000US-0206405.  
 XX (PION-) PIONEER HI-BRED INT INC.  
 PA (CURA-) CURAGEN CORP.  
 XX Bidney D., Craesta OR, Hu X, Lu G;  
 PI WPI; 2001-211215/21.  
 XX DR P-PSDB; AAB72303.

Novel isolated defence-related signalling gene isolated from sunflower encoding neoxanthin cleavage enzyme, amino acid permease or glutamic acid-rich protein useful for increasing resistance of plant to a pathogen

Claim 1; Page 94-97; 135pp; English.

This invention relates to defence-related signalling genes isolated from the sunflower (*Helianthus annuus*). The genes encode a neoxanthin cleavage enzyme (NCE), an amino acid permease (AAP) and a glutamic acid rich protein (GRP). The signalling gene is useful for increasing the resistance of a plant to a pathogen such as fungus, virus, bacterium, nematode or insect (e.g. European corn borer), preferably incorporating a *Sclerotinia* spp., *Phoma* spp., or *Phomopsis* spp., by stably incorporating a construct containing the gene into the genome of the plant. The gene is useful for regulating gene expression in a plant, in response to a stimulus such as infection with a pathogen, damage from a pathogen, hydroperoxide, jasmonic acid, methyl jasmonate, salicylic acid, oxalic acid or expression of a gene encoding oxalic acid oxidase. The genes are also useful for stem-preferred regulation of gene expression in a plant. The genes are useful in agriculture, particularly in the breeding of crop plants with improved agronomic traits, for modifying abscisic acid (ABA) metabolism and for modifying amino acid transport and content in plants. The present sequence represents cDNA encoding the sunflower neoxanthin cleavage enzyme (NCE).

Sequence 1950 BP; 535 A; 456 C; 453 G; 506 T; 0 other;

Query Match 9.4%; Score 169.2; DB 22; Length 1950;

Best Local Similarity 51.0%; Pred. No. 18-43;

Matches 480; Conservative 0; Mismatches 453; Indels 9; Gaps 3;

QY 347 CGGCTGATCCTAGTGTTCAGATCGCGGAAATTTTGTCTCGGTGAATGAACAGCCGTC 406

Db 308 CAGTTGATCCAAAACACAGTTTGTCTGATACTTTTACCGGTGACGAACCTCCCTCGA 367

QY 407 GGCGTAATCTTCGGTGTTCGGAACCTTCGGAATTCATCAAGAGTGTATGCGCA 466

Db 368 CTGACTGTGAAGTCATCGAGGCGCACTGCCAAGTTGCCCTTACGGTGCTTACTCCGTA 427

QY 467 ACGGAGCTAACCCACTTCACGAGCCGGTGACAGGTACCACTTCTTCCAGGAGACGGTA 526

Db 428 ATGTCGGAACCCGCAATTCCTTCGCGAGGACCCCTACCACTCTTCGATGCGGATGCA 487

QY 527 TGGTTACGCGCTCAATTCGHAACGTTTCAGCTAGCTACGCTTCCGCTTACTCTAGA 586

Db 488 TGCTCATGCTATTCGTAICTCAATGGAAGTCTCGTTATGTAGCCGATACATCAAAA 547

QY 587 CTAAACCGGTTTGTTCAGGAACGTCATTCGGTTCGACCGGTTTCCCAAGCATCGGTG 646

Db 548 CATACAAATATTCAATAGAGAAAGAGCGGATTCCTCCATTTATTCAAAACGTTTTCAG 607

QY 647 AGCTTCAGCGCCACACCGGTATTGCCGCTCATGCTATTC---TACGCCACAGCTGCAG 703

Db 608 GGTTAATGCTGACTGCTCTGAGCTCGCATGCGACTCACTGCCGCCGATTTTGG 667

QY 704 CCGGTATAGTCGACCCCGGACACCGGAACCGGTGTAGCTAAACCGCGGTTTGTGCTATTCA 763



```
Db 668 CTGGAAATTTGACCCCAAAAGATATTGGTCTAGCCAAATACCAGTCTGECCTTTTGG 727
Qy 764 ATGCGCGGTTATTTGGCTATGTCGGAGGATGATTTTACCTTACCAAGTTTACATCACTCCCA 823
Db 728 GCAACAGACTTTTGGCTCTGGAGAGTGGGATCTCCCATATGCGTCAAACTAGCGCCG 787
Qy 824 ATGGAGATTTAAACCGTTGGTGGTTCGATTTTGTATGGACAATTAGAATCCACAATGA 883
Db 788 ACGGTGACATATGTCACCGTTCGGAGCTTGGAGCTTGGACGGCAAACTATTATGAGCATGA 847
Qy 884 TTGCCCAACCCGAAAGTCCAGCCGGAATCCGGTGAATCTTCGCTTTAAGCTTACGACGTCG 943
Db 848 CCGCTCACCCCAAAATCCATCCAGTACGAAAGAGCTTTTGGCTTTTCGTTACGTCAG 907
Qy 944 TTTCAAAGCCTTACCTAAATACTTCCGATTTCTCACGGACGGAATTAATCACCGGACG 1003
Db 908 TCCCC--CCTTTCCTAACCTTTTTCGTTTCAACGAAACGGAGAAAAACAAGCCGATG 964
Qy 1004 TCGAGATTCAGC---TTGATCAGCCACAGGATGATGCACGATTTCCGCTTACAGAGAACT 1060
Db 965 TCCCGATCTTCTCATATGACAGCCGTCGTTTCTCCAGACTTCGCCATCACCAAAACT 1024
Qy 1061 TCGTCGTCGTACCTGACCAAGATGTCGTTTTCAGCTCCCGGAGATGATCCGCGTGGGT 1120
Db 1025 ACGCGATTTTCCCGAGATCCAAATCGGGATCAGCCCAATGGAGATGCTGGGTGGGGAT 1084
Qy 1121 CTCGGGTGTTTACACAAGAACAGGTCGCAAGATTCGGGATTTTAGACAAATACGCG 1180
Db 1085 CCCCCTTAGCGCGGACGCTGGAAGGTGCTCGGCTCGGTTGATCCCTCGGTACGGGA 1144
Qy 1181 AAGATTCATCGAACHTTAAGTGGATTGATGCTCCAGATTGCTTCTGTTCCATCTCTGGA 1240
Db 1145 AAGACGAGTCCGAGATGAAGTGGTTGAGTTCCGGGTTTAAATGATGATACATTGCCATCA 1204
Qy 1241 ACGTTGGAGAGCCAGAACAGATGAAGTCGTCGTGATAG 1282
Db 1205 ATGCATGGAGAGGATGCGGAGATACGGTGGTATGGTG 1246
```

Search completed: November 16, 2003, 15:31:21  
Job time : 515 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 16, 2003, 20:16:12 ; Search time 3372 Seconds  
(without alignments)  
4317.432 Million cell updates/sec

Title: US-09-758-269-6  
Perfect score: 3150  
Sequence: 1 MASTATAAAGSRWLGNGHT.....VPYGFHGTFIGADDLAKQVV 539

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22761392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh

-MODEL=frame+ p2n model  
-Q/cgn2\_1/USPTO spool/US09758269/runat 14112003 192309 25769/app query.fasta\_1.775  
-DE=EST -QFMT=fastap -SUFFIX=first -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=hamand40.cdi -LIST=45  
-LOCALALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09758269 @CGN 1 1.2810 @runat 14112003 192309 25769 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_hic:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_hic:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: em\_gss\_hum:\*
- 18: em\_gss\_inv:\*
- 19: em\_gss\_pin:\*
- 20: em\_gss\_vrt:\*
- 21: em\_gss\_fun:\*
- 22: em\_gss\_mam:\*
- 23: em\_gss\_mus:\*
- 24: em\_gss\_pro:\*
- 25: em\_gss\_rod:\*
- 26: em\_gss\_pg:\*
- 27: em\_gss\_vrl:\*
- 28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	1256.5	39.9	837	28	BH662445	BH662445 BOHTU09TR
C 2	1108	35.2	805	12	BM408615	BM408615 EST582942
C 3	1073	34.1	766	29	BZ457573	BZ457573 BOMHT80TF
C 4	1057	33.6	720	12	BM412731	BM412731 EST587047
C 5	1036	32.9	721	28	BH739063	BH739063 BOMHT72TF
C 6	1020	32.4	787	12	BM408565	BM408565 EST582892
C 7	1013	32.2	781	28	BH549344	BH549344 BOMHT053TF
C 8	977	31.0	644	12	BM535408	BM535408 EST588430
C 9	956	30.3	643	10	AM933524	AM933524 EST59283
C 10	936.5	29.7	1884	11	AY106323	AY106323 Zea mays
C 11	909	28.9	637	13	BU550566	BU550566 GM880021A
C 12	888	28.2	667	29	AL950790	AL950790 Arabidops
C 13	876	27.8	592	29	BF113346	BF113346 EST440536
C 14	865	27.5	627	9	AV826228	AV826228 AY826228
C 15	854	27.1	668	29	CC157378	CC157378 I919B12.b
C 16	847.5	26.9	594	28	BH458011	BH458011 BOHT00TF
C 17	846.5	26.9	696	28	BZ022957	BZ022957 oeh32C12
C 18	840	26.7	566	10	BE461924	BE461924 EST413439
C 19	826	26.2	564	12	BM085488	BM085488 saj37d09
C 20	825	26.2	553	12	BM085672	BM085672 saj38a02
C 21	821.5	26.1	711	29	BZ429431	BZ429431 EONDR14TR
C 22	817	25.9	562	12	BM536135	BM536135 EST389157
C 23	815.5	25.9	549	12	BM891057	BM891057 sam22d05
C 24	815	25.9	617	10	BE458861	BE458861 EST414153
C 25	813	25.8	649	29	CC016051	CC016051 PUDGL12TD
C 26	804.5	25.5	547	12	BI974879	BI974879 saj74b11
C 27	802.5	25.5	602	13	BE505126	BE505126 EST612541
C 28	799.5	25.4	641	29	BZ525251	BZ525251 OGALK24TM
C 29	799	25.4	559	12	BM085005	BM085005 saj3la08
C 30	798	25.3	520	10	BE432853	BE432853 EST399478
C 31	798	25.3	547	10	BE437072	BE437072 EST408190
C 32	794	25.2	627	10	BF051297	BF051297 EST436472
C 33	791	25.1	657	9	AA556214	AA556214 69 Lob101
C 34	786.5	25.0	618	12	BM084948	BM084948 saj30b08
C 35	786	25.0	546	9	AM443298	AM443298 EST308228
C 36	769	24.4	578	29	BX004482	BX004482 Arabidops
C 37	765.5	24.3	612	29	BZ525247	BZ525247 OGALK24TC
C 38	761	24.2	532	12	BU563195	BU563195 BU563195
C 39	759.5	24.1	567	13	BQ582721	BQ582721 E012280-0
C 40	749	23.8	648	29	BZ727312	BZ727312 OGFBF70TC
C 41	739	23.5	495	10	BE434930	BE434930 EST406008
C 42	734.5	23.3	509	10	BE451573	BE451573 EST402461
C 43	716.5	22.7	699	28	BH981407	BH981407 odel6e05
C 44	711	22.6	651	29	BZ462969	BZ462969 BONDN77TR
C 45	701.5	22.3	590	10	BG592987	BG592987 EST491665

# ALIGNMENTS

RESULT 1  
BH662445/c  
LOCUS BOHTU09TR BO\_2.3 KB Brassica oleracea genomic clone BOHTU09, linear GSS 19-FEB-2002  
DEFINITION BOHTU09TR BO\_2.3 KB Brassica oleracea genomic survey sequence.  
ACCESSION BH662445  
VERSION BH662445.1 GI:18721309  
KEYWORDS GSS.  
SOURCE Brassica oleracea  
ORGANISM Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; eurosids II; Brassicales; Brassicaceae; Brassica.

```

REFERENCE 1 (bases 1 to 837)
AUTHORS Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished
COMMENT Contact: Chris Town
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
FEATURES
    Location/Qualifiers
        1..837
            /organism="Brassica oleracea"
            /mol_type="genomic DNA"
            /strain="TO1000DH3"
            /db_xref="taxon:3712"
            /clone="BOHTU09"
            /clone_lib="BO.2.3 KB"
            /note="Vector: pHOs1; Site 1: BatXI; 2-3 kb sheared genomic DNA inserted into pHOs1 using BstXI linkers"
BASE COUNT 171 a 238 c 220 g 208 t
ORIGIN
Alignment Scores:
Pred. No.: 3,78e-122 Length: 837
Score: 1256.50 Matches: 238
Percent Similarity: 93.19% Conservative: 22
Best Local Similarity: 85.30% Mismatches: 17
Query Match: 39.89% Indels: 2
DB: 28 Gaps: 1
US-09-758-269-6 (1-599) x BH662445 (1-837)
QY 270 valGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGln 289
DQ 836 GTCAGGATCACTCAAGGGAGACTGATAAAACCGTTGGCCGTACGATTTCACACACAGT 777
QY 290 LeuGluSerThrMetIleAlaHisProLysValAspProGluSer-GlyGluLeuPheAl 309
DQ 776 TTAGATCCACATGATCGCCACCCGAAAGTCGACCGGAATCCGGGGAGCTATTGCG 717
QY 309 aLeuSerTyAspValValSerLysProTyLeuLysTyPheArgPheSerProAspGl 329
DQ 716 TCTAAGCTACGACGCTGTTTCGAAGCCTTTTAAAGTACTTCAGACTCTCCCGACCG 657
QY 329 yThrLysSerProAspValGluIleGlnLeuAspGlnProThrMetMetHisAspPheAl 349
DQ 656 GATAAAATCAACCGGATGTCAGATCAACCTCGATCAGCCGACGATGATGACGACTTCGC 597
QY 349 alleThrGluAsnPheValValValProAspGlnValValPheLysLeuProGluMe 369
DQ 596 TATAACGGAGAACTTCGTGTATACCGGACCAAGTCTGTGTTCAAGTCTCCGGAGAT 537
QY 369 tIleArgGlyGlySerProValValTyAspLysAsnLysValAlaArgPheGlyIleLe 389
DQ 536 GATCCGCGCGGCTCTCCGGTGTGTACGACAAAGGAGAAAGTTGCAAGATTCCGAGTTT 477
QY 389 uAspLysTyAlaGluAspSerSerAsnIleLysTrpIleAspAlaProAspCysPheCy 409
DQ 476 AGACAAATGACCGCGGACTGTCGCCATCATGCTGATGATGATGATGATGATGATGAT 417
QY 409 sPheHisLeuTrpAsnAlaTrpGluLup-cGluThrAspGluValValIleGlySe 429
DQ 416 CTTTCATCTCTGGAACGCTTGGGAGAGCCGAGACAGAGAGATCGTGTATCGGTC 357
QY 429 rCysMetThrProProAspSerIlePheAsnGluSerAspGluAsnLeuLysSerValle 449
DQ 356 GTGCATGACTCCCGCGGACTCAATTTTCAACGAGCGCGGAGAAATCTCGAGAGCTCT 297
QY 449 userGluIleArgLeuAsnLeuLysThrGlyGluSerThrArgArgProIleIleSerAs 469

```

```

Db 296 CTCCAAGATACGCTGAACCTCAGAACCGAGAGTCCACTCGCCGCTGATCATCTCGA 237
QY 469 nGluAspGlnGlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgly 489
DQ 236 CGGAGAGGAGCAAGTCAACCTCGAACGAGGATGTTTAAACCGAACATCTCGCCGTA 177
QY 489 sThrLysPheAlaTyLeuAlaLeuAlaGluProTrpProLysValSerGlyPheAlaLy 509
DQ 176 GACCAGGTTTCGCTTACCTCGCTTACCGAGCGGTGGCCCTAAAGTCTCTGGCTTTG 117
QY 509 sValAspLeuThrThrGlyGluValLysLysHisLeuTyGlyAspAsnArgTyGlyGl 529
DQ 116 AGTTGATCTCTTTACCGGAGAGTCAAGAAAGTCTGTACCGCGGTGACCGTTCGCTG 57
QY 529 yGluProLeuPheLeuProGly---GluGlyGlyGluGluAspGluGlyTyIle 546
DQ 56 AGAGCTCTCTGTTTCTCCAGGAGACAGAGACAGAGGACGACGCGGTATCATC 2
RESULT 2
BM408615 805 bp mRNA linear EST 22-JAN-2002
LOCUS EST592942 tomato breaker fruit Lycopersicon esculentum cDNA clone
DEFINITION cLEG45023 5' end, mRNA sequence.
ACCESSION BM408615
VERSION BM408615.1 GI:18260245
KEYWORDS EST.
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 805)
AUTHORS Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Teai
,J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S., Ronning
,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue, breaker stage (2002)
Unpublished
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
This clone is available through the Clemson University Genomics
Institute
Seq primer: T3.
FEATURES
    Location/Qualifiers
        1..805
            /organism="Lycopersicon esculentum"
            /mol_type="mRNA"
            /cultivar="TA496"
            /db_xref="taxon:4081"
            /cldes="cLEG45023"
            /tissue_type="Pericarp"
            /dev_stage="breaker"
            /lab_host="SOLR"
            /clone_lib="tomato breaker fruit"
            /note="Vector: pBluescriptKCMuadapt; Site 1: EcoRI;
            Site 2: XhoI; supplier: Boyce Thompson Institute;
            Sequencing: The Institute for Genomic Research. Fruit
            were harvested at the breaker stage (first sign of
            lycopene accumulation on the blossom end of fruit). Fruit
            were cut in half and the seeds and locules were discarded
            prior to freezing the pericarp."
BASE COUNT 262 a 144 c 180 g 219 t
ORIGIN
Alignment Scores:
Pred. No.: 1.82e-106 Length: 805
Score: 1108.00 Matches: 212
Percent Similarity: 89.51% Conservative: 27
Best Local Similarity: 79.40% Mismatches: 26
Query Match: 35.17% Indels: 4

```

DB:	12	Gaps:	1
US-09-758-269-6 (1-599) x BM408615 (1-805)			
QY	300	ValAspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerIysLeuProTyr	319
Db	5	CTCGACCCAGCTTTCCGGTGAGCTATTGCTCTTAGCTAGCTAGTGATTCAGAGCCATAC	64
QY	320	LeuLysTyrPheAArgPheSerProAspGlyThrLysSerProAspValGluLeuLeu	339
Db	65	CTCAAGTACTTCAGATTTTCAAAAAATGGGGAAAAATCAAATCATGTTGAAATTCAGATT	124
QY	340	AspGlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAsp	359
Db	125	GAAGACCCAAACAAATGATGATTTCCGATTTACTGAGAACTTCGTCGTCATTCCTCAT	184
QY	360	GlnGlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyrAsp	379
Db	185	CAACAAGTCGTTTCAAGATCTGAAATGATCCGTGGAGGTTCACCGGTGGTTTACGAC	244
QY	380	LysAsnLysValAlaAArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIle	399
Db	245	AAGAACAAGTTTCCCGATTGGTATTCTGGTAAGTACGGAAGATGGGTCTCATTTTG	304
QY	400	LysTrpIleAspAlaProAspCysPheHisLeuTrpAsnAlaTrpGluGluPro	419
Db	305	AAATGGGTGAAGTACCTGATTGTTCTGTTCCACTCTGGAATGCTTGGGAAGAACA	364
QY	420	GluThrAspGluValValIleGlySerCysMetThrProProAspSerIlePheAsn	439
Db	365	GAACACAGATGAATCGTTGTAATTTGGTTTCATGTATGACACCACTCCACTTTTCAAT	424
QY	440	GluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGly	459
Db	425	GAATGTGATGAAGGCTAAAGAGTGTTTATTCGAAATCCGTCCTCAATTTGAAACAGG	484
QY	460	GluSerThrArgArgProIleIleSerAsnGluAspGlnValAsnLeuGluAlaGly	479
Db	485	AAATCAACAAGAAAAATCCATTAATCGAAACCCGGATGAAACAAGTGAAATTTAGAAGCTGA	544
QY	480	MetValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTrpLeuAlaLeuAlaGlu	499
Db	545	ATGGTGAAACCGAACAACAACTCGGAAGAAACAGAGTATGCTTATTTGGCTATCGCTGAA	604
QY	500	ProTrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLys	519
Db	605	CCATGGCCAAAAGTTTCTGGTTTTCGAAAAGTAAACCTGTTCAACCGGTCAAGTTTGAGAAA	664
QY	520	HisLeuTyrGlyAspAsnArgTyrGlyGluProLeuPheLeuProGlyGlu-----	537
Db	665	TTCAATTATGGTGCACAAACAAATATGGTGGGGAACCTCTTTTTTACAAGA-GAACCCCAAC	723
QY	538	GlyGlyGluGluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLys	557
Db	724	AGCCAGGAAGAGACGATGGTTATATTAGCTTCGTTCCATCGATGAGAAAGA-TGGGAA	782
QY	558	SerGluLeuGlnIleValAsn	564
Db	783	TCAAAACCTGCAAAATGGTAAAC	803

RESULT 3	B2457573	766 bp	DNA	linear	GSS 13-DEC-2002
BZ457573/c	BONG80TF BO_1.6_2_KB_tot				
LOCUS	BONG80TF BO_1.6_2_KB_tot				
DEFINITION	Brassica oleracea genomic clone BONG80, genomic survey sequence.				
ACCESSION	BZ457573				
VERSION	BZ457573.1	GI:26737071			
KEYWORDS	GSS.				
SOURCE	Brassica oleracea				
ORGANISM	Brassica oleracea				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; eurosids II; Brassicales; Brassicaceae; Brassica.				
	; eurosids II; Brassicales; Brassicaceae; Brassica.				

QY 353 nPheValValProAspGlnGlnValValPheLysLeuProGluMetIleArgGlyG1 373  
 Db 220 CTTCTGCTGTTTCCGGACCGACGAGTGTGTTTCCAGGCTACCGAGATGATCCGCGCG 161  
 QY 373 ySerProValValTyAspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyAl 393  
 Db 160 CTCTCCGGTGGTTTACGACAAACAAAGAAATCAAGATTCGGTGTGTTTGAGCAATACGC 101  
 QY 393 aGluAspSerSerAenIleLysTyrIleAspAlaProAspCysPheCysPheHisLeuTr 413  
 Db 100 TGAAGACCGTTCGTGATCGCGGTGATCGAGTGAAGACTGCTTCGTTTCCATCTCTG 41  
 QY 413 pAenAlaTrpGluGluProGluThrAspGluValVal 426  
 Db 40 GAACGCTTGGGAAGAGCGGACACAGAGGATCGTCGTC 1

## RESULT 4

BM412731 720 bp mRNA linear EST 22-JAN-2002  
 LOCUS EST587047 tomato breaker fruit Lycopersicon esculentum cDNA clone  
 DEFINITION CLEGG60N24 5' end, mRNA sequence.

## ACCESSION

BM412731

## VERSION

BM412731.1 GI:18264350

## KEYWORDS

EST.

## SOURCE

Lycopersicon esculentum (tomato)

## ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

1 (bases 1 to 720)

## REFERENCE

Alcala,J., Vrabalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai

,J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S., Ronning

,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.

Generation of ESTs from tomato fruit tissue, breaker stage (2002)

Unpublished

Contact: CUGI

Clemson University Genomics Institute

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

This clone is available through the Clemson University Genomics

Institute

Seq primer: T3.

Location/Qualifiers

1..720

/organism="Lycopersicon esculentum"

/mol\_type="RNA"

/cultivar="TA96"

/db\_xref="taxon:4081"

/clone="CLEGG60N24"

/tissue\_type="Pericarp"

/dev\_stage="breaker"

/lab\_host="SOLR"

/clone\_lib="tomato breaker fruit"

/note="vector: BluescriptSKmCudapt; Site 1: EcoRI;

Site 2: XhoI; supplier: Boyce Thompson Institute. Fruit

sequencing: The Institute for Genomic Research. Fruit

were harvested at the breaker stage (first sign of

lycopene accumulation on the blossom end of fruit). Fruit

were cut in half and the seeds and locules were discarded

prior to freezing the pericarp."

BASE COUNT 226 a 143 c 155 g 196 t

## ORIGIN

Alignment Scores:

Pred. No.: 3,84e-101 Length: 720

Score: 1057.00 Matches: 194

Percent Similarity: 90.38% Conservative: 22

Best Local Similarity: 81.17% Mismatches: 23

Query Match: 33.56% Indels: 0

DB: 12 Gaps: 0

US-09-758-269-6 (1-599) x BM412731 (1-720)

QY 248 AsnAlaGlyLeuValTyPheAsnGlyArgIleuLeuAlaMetSerGluAspLeuPro 267  
 Db 3 AACCGCGGTTTAACTCTAATTAACCGATTACTTGTCTGTAAGATGATTGGCT 62  
 QY 268 TyrGlnValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAsp 287  
 Db 63 TACCATTGTAAGAGTAACACCCCGCGGATCTTAAACAGAGGGTCGATTCGATTTCGAC 122  
 QY 288 GlyGlnLeuGluSerThrMetIleAlaHisProLysValAspProGluSerGlyGluLeu 307  
 Db 123 GGCCAGCTAAATATCCACCATGATAGCTACCCAAAGCTCGACCCAGTTTCCGGTGAGCTA 182  
 QY 308 PheAlaLeuSerTyAspValValSerLysProTyLeuLysTyPheArgPheSerPro 327  
 Db 183 TTTGCTCTTACGATGATGATTCGAAGCCATACCTCAAGTACTTCAGATTTTCAAAA 242  
 QY 328 AspGlyThrLysSerProAspValGluIleGlnLeuAspGlnProThrMetHisAsp 347  
 Db 243 AATGGGGAATAATCAANTGATGTTGAATTCAGTTGAAGACCCCAACATGATGATGAT 302  
 QY 348 PheAlaIleThrGluAsnPheValValProAspGlnGlnValValPheLysLeuPro 367  
 Db 303 TTCGCAATTACTGAGAACTTCGTCGTCATCTCTGATCAACAAAGTCGTTTTCAGATGCT 362  
 QY 368 GluMetIleArgGlyGlySerProValValTyAspLysAsnLysValAlaArgPheGly 387  
 Db 363 GAATGATCCGTGAGGTTTCCCGGTGTTTACGACAGACAAAGTTTCCCGATTGGT 422  
 QY 388 IleLeuAspLysTyAlaGluAspSerSerAsnIleLysTrpIleAspAlaProAspCys 407  
 Db 423 ATTCTGATTAAGTACCGAAGATGGTCTGATTTGAAATGGGTGGAAGTACCTGATTGT 482  
 QY 408 PheCysPheHisLeuTrpAenAlaTrpGluGluProGluThrAspGluValValIle 427  
 Db 483 TTCTGTTTCCACTCTGGATGCTTGGGAAGACGAAACACATGAATCGTTGTAAT 542  
 QY 428 GlySerCysMetThrProAspSerIlePheAsnGluSerAspGluAsnLeuLysSer 447  
 Db 543 GGTTTCATGTATGACACACAGACTCCATTTTCAATGAATGTGATGAAGGGCTAAAGAT 602  
 QY 448 ValLeuSerGluIleArgLeuAsnLeuLysThrGlyGluSerThrArgArgProIleIle 467  
 Db 603 GTTTTATCCGAATCCGTCCTCAATTTTGAACAGGGAATTCACAAAGAAATCCATAATC 662  
 QY 468 SerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeu 486  
 Db 663 GAAACCCCGGATGAACAAAGTGGATTTAGAAGCTGGAATGGTGAACCCGAAACAACTC 719

## RESULT 5

BM412731

## LOCUS

BM412731

## DEFINITION

721 bp DNA linear GSS 20-FEB-2002

BOMHT72TF BO\_2\_3 KB Brassica oleracea genomic clone BOMHT72,

genomic survey sequence.

ACCESSION

BH739063

VERSION

BH739063.1 GI:18844458

KEYWORDS

GSS

SOURCE

Brassica oleracea

Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; euroids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 721)

REFERENCE

Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished

Other\_GSSs: BOMHT72TR

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: [cdtown@tigr.org](mailto:cdtown@tigr.org)

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

# FEATURES

Source

Location/Qualifiers  
1..721  
/organism="Brassica oleracea"  
/mol\_type="genomic DNA"  
/strain="TO1000DH3"  
/db\_xref="taxon:3712"  
/clone="BOMHT72"  
/clone\_lib="BO 2.3 KB"  
/note="vector: pHO51, Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHO51 using BstXI linkers"

BASE COUNT 188 a 178 c 198 g 157 t

## ORIGIN

### Alignment Scores:

Pred. No.: 6.4e-99 Length: 721  
Score: 1036.00 Matches: 193  
Percent Similarity: 91.56% Conservative: 24  
Best Local Similarity: 81.43% Mismatches: 16  
Query Match: 32.89% Indels: 4  
DB: 28 Gaps: 2

US-09-758-269-6 (1-599) x BH739063 (1-721)

QY 302 ProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeuLys 321  
DB 4 CCGGAACCCCGAGCTATTGCTCTACGTCACGACGTCGTTTCGAGCCTTACTTAAAA 63  
QY 322 TyrPheArgPheSerProAspGlyThrLysSerProAspValGluLeuLysGln 341  
DB 64 TACTTCAGATTCGCATCGACGCGTTCGCAATCCCGACGTCGAGATCCCGCTCGACAA 123  
QY 342 ProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAspGlnGln 361  
DB 124 CCGAGATGATCCAGCATTCGCGATCCGAGAACCTTCGCGGTTCGCGACGACGAA 183  
QY 362 ValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyrAspLysAsn 381  
DB 184 GTGGTGTTCAGGCTACCGAGATGATCCGCGCGCTCTCCGGTGGTTTACGACAAAC 243  
QY 382 LysValAlaArgPheGlyLeuAspLysTyrAlaGluAspSerSerAsnIleLysTrp 401  
DB 244 AGAATCAAGATTCGGTGTGTTGAGCAATACGCTGAAGACGCTTCGTCATCGCGTGG 303  
QY 402 IleAspAlaProAspCysPheHisLeuTyrAsnAlaTyrGluLysProGluThr 421  
DB 304 ATCGAGTGGAGATGCTTCTGTTTCCATCTCTGGAACGCTTGGGAAGCGCGAGACA 363  
QY 422 AspGluValValIleGlySerCysMetThrProProAspSerIlePheAsnGluSer 441  
DB 364 GACGAGGTGCTGTCATCGGTCATGATGACGCGCGCGCTCGATATTCAACGACAC 423  
QY 442 AspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGluSer 461  
DB 424 GACGAAACACTTCAGAGTGTGTTGCGAGATPAGACTAACCTGAAACAGGGAGTCC 483  
QY 462 ThrArgArgProIleIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMetVal 481  
DB 484 ACGCGTGGCGGTTATCTCC-----GAGAAATTAATCTCGAAGCGGTATGGTA 534  
QY 482 AsnArgAsnMetLeuGlyArgGlyThrLysPheAlaTyrLeuAlaLeuAlaGluProTrp 501  
DB 535 ACCCGGAATCTTTTAGTAGGAGAAAACCGCGTTCGCTTACCTTGCCTTTAACCGACCGTGG 594  
QY 502 ProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHisLeu 521  
DB 595 CCTAAGTGTCCGGTTTCGGAAGTGGACTTATCTACCGGAGAGAGTTTCAAAATATATA 654  
QY 522 TyrGlyAspAsnArgTyrGlyGluProLeuPheLeuProGlyGluGly 538  
DB 655 TAGCGTGACGTAAGTTCGAGGAGAGCCTCTGGTTATGCCG---GAAGGG 702

## RESULT 6

BM408565

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT

INSTITUTE

ADDRESS

TELEPHONE

FAX

E-MAIL

URL

PROJECT

PI

SPONSOR

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

REMARKS

```

Db      167 ATTTCTGGGAATTTTGTCCGGTACCGGAAATCCAGTGTGTCAATCTCTTCGGGTACC 226
Qy      143 GlyLeuProAspSerIleYsGlyValTyrValArgAsnGlyValAsnProLeuHis 162
Db      227 GGAAATAATACCAATGTCTTCAAGCGGTTTACGTTTCAAGACGAGCTTAACCTCTTTT 286
Qy      163 GluProValThrGlyHisHisPhePheAspGlyAspGlyMetValHisAlaValysPhe 182
Db      287 GAACCAACCGCGGACACCATTTCTCGACGCGGAGCGGTATGTTTCACGCGGTTCAATTC 346
Qy      183 GluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheValGlnGlu 202
Db      347 AAAATGGGTGGCTAGTTTACCGTTTACCGTTTCACTGAACACGAGGCTGTTTCAAGAA 406
Qy      203 ArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHisThrGly 222
Db      407 AAAGCTTTGGTTCGCTCTTTCCCTAAAGCCATTTGTTGAATTTACATGTTCACTCTGGA 466
Qy      223 IleAlaArgLeuValLeuPheTyrAlaArgAlaAlaGlyIleValAspProAlaHis 242
Db      467 ATTGCAAGGCTTATGCTGTTTACGCTCGTGGGCTCTTCGGACTTGTGATCACAGTAAA 526
Qy      243 GlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAlaMetSer 262
Db      527 GGAAGTGGTGTTCGAAACCGCGGTTTAGTCTATTTCAATTAACGATTACTTGTATGCT 586
Qy      263 GluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuIysThrValGly 282
Db      587 GAAGATGATTGCTTACCATTGTAAGTAAAGTAAACACCGCGGCTCTTAAACACAGGGT 646
Qy      283 ArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysValAspPro 302
Db      647 CGATTGATTTCCAGCGCCAGCTTAAATTCACCATGATAGTCAACCAAGCTCGACCA 706
Qy      303 GluSerGlyLeuLeuPheAlaLeuSerTyrAspValValSerIysPheProTyrLeuIysTyr 322
Db      707 GTTTCGGTGAGCTATTGCTTAGTACGATGATGATGATGATGATGATGATGATGATGAT 766
Qy      323 PheArgPheSerProAspGly 329
Db      767 TTCAGATTTTCAAAATGGGG 787

RESULT 7
BH549344      781 bp DNA linear GSS 14-DEC-2001
LOCUS        BOGRQ53TF BOGR Brassica oleracea genomic clone BOGRQ53, genomic
DEFINITION   survey sequence.
ACCESSION    BH549344
VERSION      BH549344.1 GI:17801124
KEYWORDS     GSS.
SOURCE       Brassica oleracea
ORGANISM     Brassica oleracea
REFERENCE    1 (bases 1 to 781)
AUTHORS      Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
TITLE        Whole genome shotgun sequencing of Brassica oleracea
JOURNAL      Unpublished
COMMENT      Other GSSs: BOGRQ53TR
              Contact: Chris Town
              TIGR
              9712 Medical Center Drive, Rockville, MD 20850, USA.
              Tel: 301-838-3523
              Fax: 301-838-0208
              Email: cdtown@tigr.org
              DNA is from a doubled haploid provided by Tom Osborn.
              Seq primer: TF
              Class: sheared ends.
              Location/Qualifiers
                1..781
                  /organism="Brassica oleracea"
                  /mol_type="genomic DNA"

```

```

/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGRQ53"
/clone_lib="BOGR"
/note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
       Genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT  204 a 203 c 192 g 182 t
ORIGIN
Alignment Scores:
Pred. No.: 1.95e-96 Length: 781
Score: 1013.00 Matches: 190
Percent Similarity: 84.23% Conservative: 29
Best Local Similarity: 73.08% Mismatches: 41
Query Match: 32.16% Indels: 0
DB: 28 Gaps: 0
US-09-758-269-6 (1-599) x BH549344 (1-781)
Qy      92 AlaAlaAlaAlaAlaLeuAspAlaAlaGlyPheLeuValSerHisGlyLysLeuHis 111
Db      1 GCAGCGCGGATTGCGATCGATGATGCTGAGCGAGCGTTGATCTCAGCTGAGCAAGATACT 60
Qy      112 ProLeuProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAlaProValAsn 131
Db      61 CTTCTTCCCAAAACCGCTGATCCAGCTATTCAAATCCGCGGAATTTATTTCCCGGTGCGG 120
Qy      132 GluGlnProValArgAsnLeuProValValGlyLysLeuProAspSerIleLysGly 151
Db      121 GAATTTCCGCTCGAGAGTCTCTCGACGTTGAAGGAAACATCCCTGACTGCATTACGGA 180
Qy      152 ValTyrValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHisHisPhePhe 171
Db      181 GTTATGCTCCGTAAACCGTCAATTCGATGTTCCAGCAATCTGCTGGCATCATTTGTTTC 240
Qy      172 AspGlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerTyrAlaCys 191
Db      241 GACGAGACCGGAATGGTTTCACGAGTTTAAATACCAACCGGTTCCAGCAAGCTACGCAATGC 300
Qy      192 ArgPheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPhePro 211
Db      301 CGGTTTACTAAACCGAGAGATTAAATTCAGAAAAACGATTGGTTCGACCGGTTTTTCG 360
Qy      212 LysAlaIleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyrAla 231
Db      361 AAAGCAATTCGCGAGCTTCACGCCACTCGGGAATCGACGTCGTGATGCTTTTTTACGCA 420
Qy      232 ArgAlaAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAlaAsnAlaGlyLeu 251
Db      421 CGTGGGCTTTCGGCGCTCGTCAACATCAAAACCGGCTCGGGTAGCGAAGCGCGGTTG 480
Qy      252 ValTyrPheAsnGlyArgLeuLeuAlaMetSerGluAspAspLeuProTyrGlnValGln 271
Db      481 GTTTACTTCAATAACCGGCTTTTAGCAATGTTCAGAGACCGATTATTCGTTACCAATTAATA 540
Qy      272 IleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeuGlu 291
Db      541 ATAATCAAAACCGGCGACCTCCAAACCGTTGGGCTTACGATTTCCGCGGTCACTTAAAA 600
Qy      292 SerThrMetIleAlaHisProLysValAspProGluSerGlyGluLeuPheAlaLeuSer 311
Db      601 TCATCGATGATCGCCACCCGAAACTCGACCCGGTTTAAAGAGAGCTCCACGCGCTAAGC 660
Qy      312 TyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerProAspGlyThrLys 331
Db      661 TACGACGCTGTTAAAGACCTTACCTGAAATTTTTCAGATTTCTCGCGGACGCGGTAAA 720
Qy      332 SerProAspValGluIleGlnLeuAspGlnProThrMetHisAspPheAlaIleThr 351
Db      721 TCACCGGAGCTGAGATCCCGCTCGAGACTCCGACGATGGTTTCAGATTTTCGTTATACG 780

```

RESULT 8  
BMS35408



```

LOCUS      BM535408                644 bp    mRNA    linear    EST 20-FEB-2002
DEFINITION EST588430 tomato breaker fruit Lycopersicon esculentum cDNA clone
cLEGE5N24 5' end, mRNA sequence.
ACCESSION  BM535408
VERSION     BM535408.1  GI:18814035
KEYWORDS   EST.
SOURCE     Lycopersicon esculentum (tomato)
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE  1 (bases 1 to 644)
AUTHORS    Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai
J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S., Ronning
C.M., Frazer,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue, breaker stage (2002)
Unpublished
CONTACT    Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
This clone is available through the Clemson University Genomics
Institute
Seq primer: T3.
FEATURES   Location/Qualifiers
            1..644
             /organism="Lycopersicon esculentum"
             /mol_type="mRNA"
             /cultivar="TA496"
             /db_xref="taxon:4081"
             /clone="cLEGE5N24"
             /tissue_type="Pericarp"
             /dev_stage="breaker"
             /lab_host="SOLR"
             /clone_lib="tomato breaker fruit"
             /note="Vector: pBluescriptSKmCtadapt; Site 1: EcoRI;
             Site 2: XhoI; supplier: Boyce Thompson Institute; Fruit
             sequencing: The Institute for Genomic Research. Fruit
             were harvested at the breaker stage (first sign of
             lycopenic accumulation on the blossom end of fruit). Fruit
             were cut in half and the seeds and locules were discarded
             prior to freezing the pericarp."
BASE COUNT 195 a 129 c 138 g 182 t
ORIGIN
Alignment Scores:
Pred. No.: 9,44e-93    Length: 644
Score: 977.00    Matches: 178
Percent Similarity: 91.59%    Conservative: 18
Best local Similarity: 83.18%    Mismatches: 18
Query Match: 31.02%    Indels: 0
DB: 12    Gaps: 0

US-09-758-269-6 (1-599) x BM535408 (1-644)

Qy      248  AsnAlaGlyLeuValTyrPheAnGlyArgLeuLeuAlaMetSerGluAspLeuPro 267
Db      3   AACGCCGGTTAGCTATTCTCAATACCCGATTCCTGCTATGCTGAGATGATTGGCT 62
Qy      268  TyrGlnValGlnIleThrProAnGlyAspLeuLeuThrValGlyArgPheAspPheAsp 287
Db      63  TACCATGTAAGGTAACACCCACCGCGGATCTTAAACACAGAGGGTTCGATTCGATTCGAC 122
Qy      288  GlyGlnLeuGluSerThrMetIleAlaHisProLysValAspProGluSerGlyGluLeu 307
Db      123  GGCACGTAATATCCACCATGATAGCTACCCAAAGCTCGACCCAGTTCCGGTGAGCTA 182
Qy      308  PheAlaLeuSerTyrAspValValserLysProTyrLeuLysTyrPheArgPheSerPro 327
Db      183  TTTCGCTTAGCTACGATGATTCAGAGCCATACCTCAAGTACTTCAGATTTCAAAA 242
Qy      328  AspGlyThrLysSerProAspValGluIleGlnLeuAspGlnProThrMetMetHisAsp 347

```

```

Db      243  ANTGGGAAAATCAATGATGTTGAAATTCAGTTGAAGACCCACATGATGATGAT 302
Qy      348  PheAlaIleThrGluAsnPheValValProAspGlnValValPheLysLeuPro 367
Db      303  TTCCGAATTACTGAGAACTTCGTCGTCATTCCTGATCAACAAGTCGTTTTCAGATGCT 362
Qy      368  GluMetIleArgGlyGlySerProValValTyrAspLysAsnLysValAlaArgPheGly 387
Db      363  GAAATGATCGTGGAGTTACCGGTGGTTTACGACAGAAACAAGATTTCCCGATTGGT 422
Qy      388  IleLeuAspLysTyrAlaGluAspSerSerAsnIleLysTyrIleAspAlaProAspCys 407
Db      423  ATTCTCGATAAGTACGGAAGATGGTCTGATTTGAAATGGTTCAGATACCTGATGTT 482
Qy      408  PheCysPheHisLeuThrAsnAlaTyrGluGluProGluThrAspGluValValIle 427
Db      483  TTCTGTTTCCACCTCTGGAAATGCTTGGGAAGACGAGAAACAGATGAATCTGTTGTAAT 542
Qy      428  GlySerCysMetThrProProAspSerIlePheAsnGluSerAspGluAsnLysSer 447
Db      543  GGTTCATGATGACACCCAGACTCCATTTCAATGAATGTGATGAGGGCTAAAGAGT 602
Qy      448  ValLeuSerGluIleArgLeuAsnLysThrGlyGluSer 461
Db      603  GTTTTATCCGAAATCCGCTCTCAATTTGAAAACGGGAATAAC 644

RESULT 9
AW933524                643 bp    mRNA    linear    EST 18-MAY-2001
LOCUS      EST39283 tomato fruit mature green, TAMU Lycopersicon esculentum
DEFINITION cDNA clone cLEF54E2 5', mRNA sequence.
ACCESSION  AW933524
VERSION     AW933524.2  GI:11388127
KEYWORDS   EST.
SOURCE     Lycopersicon esculentum (tomato)
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE  1 (bases 1 to 643)
AUTHORS    Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.B., Liang,F.,
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M.,
Frazer,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue
Unpublished
On May 30, 2000 this sequence version replaced gi:8108841.
CONTACT    CUGI
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES   Location/Qualifiers
            1..643
             /organism="Lycopersicon esculentum"
             /mol_type="mRNA"
             /cultivar="TA496"
             /db_xref="taxon:4081"
             /clone="cLEF54E2"
             /tissue_type="fruit pericarp"
             /dev_stage="mature green (3-5 days pre-ripening)"
             /lab_host="SOLR"
             /clone_lib="tomato fruit mature green, TAMU"
             /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
             XhoI; cLEF - Fruit were tagged at the 1cm stage and
             harvested 3-5 days prior to ripening. Fruit were cut in
             half to verify the seeds were indeed 'immature' and the
             seeds and locules were discarded prior to freezing the
             pericarp"
BASE COUNT 184 a 130 c 142 g 186 t 1 others
ORIGIN

```

Alignment Scores: 1.57e-90 Length: 643  
 Pred. No.: 956.00 Matches: 173  
 Score: 956.00 Conservative: 19  
 Percent Similarity: 90.14% Mismatches: 21  
 Best Local Similarity: 81.22% Indels: 0  
 Query Match: 30.35% Gaps: 0  
 DB: 10

US-09-758-269-6 (1-599) x AW933524 (1-643)

QY 230 TyrAlaArgAlaAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsnAla 249  
 Db 4 TAGCTCGTGGGCTCTTCGGACTTGTGATCAAGTAAGGAACTGGTGTTCACACGCC 63

QY 250 GlyLeuValTyrPheAsnGlyArgLeuLeuAlaMetSerGluAspAspLeuProTyrGln 269  
 Db 64 GGTTAGTCTATTTCATAACCGATTACTTGTCTGATGCTGAAGATGATTGCTTACCAT 123

QY 270 ValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGln 289  
 Db 124 GTAAGGTAACACCCACCGCGATCTTAAACAGAGGTCGATTCGATTTCCAGGCCAG 183

QY 290 LeuGluSerThrMetIleAlaHisProLysValAspProLysSerGlyGluLeuPheAla 309  
 Db 184 CTAATATCCACCATGATAGCTACCCAAAGCTCGACCCAGTTTCGGTGAGCTATTGCT 243

QY 310 LeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerProAspGly 329  
 Db 244 CTAGCTACCATGATGATTCAGAGCCATACCTCAAGTACTTCAGATTTTCACAAAATGG 303

QY 330 ThrLysSerProAspValGluIleGlnLeuAspGlnProThrMetMetHisAspPheAla 349  
 Db 304 GAAAAATCAATGATGTTGAAATCCAGTTGAGAGCCCAACATGATGATGATTTTCGCA 363

QY 350 IleThrGluAsnPheValValProAspGlnGlnValValPheLysLeuProGluMet 369  
 Db 364 ATTAATGAGAACTTCGTCGATTCATTCAGTCAACAGTCGTTTTCAGATGCTGAAATG 423

QY 370 IleArgGlyGlySerProValValTyrAspLysAsnLysValAlaArgPheGlyIleLeu 389  
 Db 424 ATCCGTGGAGTTTACCGGTGTTTACGACAGAACAAAGTTTCCCGATTTGATTTCTG 483

QY 390 AspLysTyrAlaGluAspSerSerAsnIleLysTrpLeuAspLapAspCysPheCys 409  
 Db 484 GATAGTACGCGAAGATGGGTCTGATTTGAATGGGTGAAATGATGATGATTTCTGT 543

QY 410 PheHisLeuTrpAsnAlaTrpGluProGluThrAspGluValValIleGlySer 429  
 Db 544 TTCACCTCTGGAATGCTTGGAGAGACGACAGATGAAATCGTTGTAATTTGTTCA 603

QY 430 CysMetThrProProAspSerIlePheAsnGluSerAsp 442  
 Db 604 TGATGACACCAACAGACTCCATTTTCAATGATGTGAT 642

RESULT 10  
 AY106323 1884 bp mRNA linear HTC 16-OCT-2002

LOCUS  
 DEFINITION Zea mays PC0084517 mRNA sequence.  
 ACCESSION AY106323  
 VERSION AY106323.1 GI:21209401  
 KEYWORDS HTC.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 1884)  
 Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S.,  
 Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.  
 Maize Mapping Project/DuPont Consensus Sequences for Design of  
 Overgo Probes  
 Unpublished (2002)  
 JOURNAL  
 REFERENCE 2 (bases 1 to 1884)

Authors: Coe, E.H.  
 Direct Submission  
 Submitted (25-APR-2002) Maize Mapping Project, University of  
 Missouri, Columbia, MO 65211, USA  
 If you are interested in getting corresponding physical clones,  
 these are publicly available from ZmDB and may be found by BLAST  
 searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,  
 www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the  
 maize cDNA sequences is either Virginia Walbot, Stanford or Pat  
 Schnable, Iowa State, then clones may be requested from ZmDB:  
 www.zmdb.iastate.edu.  
 Location/Qualifiers  
 1..1884  
 /organism="Zea mays"  
 /mol\_type="mRNA"  
 /db\_xref="MaizeDB:634734"  
 /db\_xref="taxon:4577"  
 /clone\_lib="Maize Mapping Project/DuPont Consensus  
 Library"  
 /note="this sequence is part of a project of EST  
 assemblies resulting from the application of public  
 contigs to seed DuPont contigs; this resource was  
 assembled by DuPont as part of a collaboration for the  
 overgo addressing of BACs in conjunction with the Maize  
 Mapping Project"

BASE COUNT 505 a 402 c 477 g 500 t  
 ORIGIN

Alignment Scores: 8.62e-88 Length: 1884  
 Pred. No.: 936.50 Matches: 211  
 Score: 936.50 Conservative: 99  
 Percent Similarity: 55.16% Mismatches: 195  
 Best Local Similarity: 37.54% Indels: 57  
 Query Match: 29.73% Gaps: 15  
 DB: 11

US-09-758-269-6 (1-599) x AY106323 (1-1884)

QY 66 SerSerAsnSerProAlaIleValValLysProLysAlaLysGluSerAsnThrLysGln 85  
 Db 94 AGCCACCAAGACGACGGCTGCTGTGTCACGCGCGCGCGCGT----- 141

QY 86 MetAsnLeuPheGlnArgAlaAlaAlaAlaAlaLeuAspAlaAlaGluGlyPheLeuVal 105  
 Db 142 -----AAGGGCTGCTTCTCTGGCGCTCGACCTGCTTGTAGTCCCTCGCGGTG 189

QY 106 -----SerHisGluLysLeuHisProLeuProLysThrAlaAspProSerValGlnIle 123  
 Db 190 CGCTCGGCCACGACAGACCAAGCGCTCCAC-----TGGCTC 228

QY 124 AlaGlyAsnPheAlaPro---ValAsnGluGlnProValArgArgAsnLeuProValVal 142  
 Db 229 TCGGCACTTCGCGCGCTGCTCGAGGAGACCGCGCGCCCAACCTTACCGTCGCG 288

QY 143 GlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnProLeuHis 162  
 Db 289 GGACACCTCCCGAGTGTGTAATGAGAGTGTGTCAGGGTTGGGCTTAATCCGAAGTTT 348

QY 163 GluProValThrGlyHisPhePheAspGlyAspGlyMetValHisAlaValLysPhe 182  
 Db 349 GCTCCTGTTGCGGGTATCCTGTTTGTATGAGAGCGGATGATTCATCGCATGCGTATT 408

QY 183 GluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheValGlnGlu 202  
 Db 409 AAGGATGGAAGAAAGTACCTATGTATCAAGATATGTGAAGACTGCCCGCTCAACAGAG 468

QY 203 ArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHisThrGly 222  
 Db 469 GAGTATTTGGTGGACCAAGTTTATGAAG---ATTGGAGACCTTAAGGATTTTGGGA 525

QY 223 IleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAspProAlaHis 242  
 Db 526 TTGTTTATGCTCCAAATGACGCAACTTCGAAAAAATTCAAAGTCTTGTGATTTCCTAT 585



University as part of the Public EST project, <http://129.186.26.94/soybeanest.html>. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, <http://web.ahc.umn.edu/biodata/nfs/soy/>. Reracking and 3' sequencing were conducted by services of the University of Illinois Kerk Center for Comparative and Functional Genomics <http://www.biotech.uiuc.edu/keck.htm>. Note: The corresponding 5' EST from each clone in the Gm-r1088 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."

BASE COUNT 160 a 180 c 126 g 171 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1,45e-85 Length: 637  
Score: 909.00 Matches: 168  
Percent Similarity: 92.38% Conservative: 26  
Best Local Similarity: 80.00% Mismatches: 14  
Query Match: 28.86% Indels: 2  
DB: 13 Gaps: 2

US-09-758-269-6 (1-599) x BU550566 (1-637)

QY 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyAlaGluAspSerAsnIleLys 400  
DB 636 AACAGGTTCCAGATTCGGATTCTCGAACAGATGCTAAAGATGCAATATGATAA 577  
QY 401 TptLAsAlaProAspCysPheCysPheHisLeuTptAsAlaTptCtLcLcProGlu 420  
DB 576 TGGATCGATGCTCCCGAGTCTTCTGTTCCACCTCTGGAACGCTGGAGAGCGGAA 517  
QY 421 ThrAspGluValValIleGlySerCysMetThrProAspSerIlePheAsnGlu 440  
DB 516 AATGATGAGATCGTTGTCTCGTTCTGATGACCCCGCGACTCCATTTTCAACGAA 457  
QY 441 SerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGlu 460  
DB 456 TGGCAAGAGAGTTGAAGAGCATCTTCTCGAGATTAAGCTGAATTTGAAGCAGCGAAG 397  
QY 461 SerThrArgArgProIleIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMet 480  
DB 396 TCCACGAGAAACCCATTATCTCG---GAATCCGACACAGTCACTTGAACCGCGATG 340  
QY 481 ValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluPro 500  
DB 339 GTCAACAGAAACAAAGCTCGGAGAAACAGCAAGTTCGGTACTTAGCACTTCGGAGCGC 280  
QY 501 TptProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHis 520  
DB 279 TGGCCTAAGGTTTCGGGTTTCCAGAGTTGATTTGTTTCACTGGGAGGTTAAGAGTAC 220  
QY 521 LeuTyrGlyAspAsnArgTyrGlyGlyGluProLeuPheLeuPro---GlyGluGlyGly 539  
DB 219 ATGATATCGGAGAGAGAGTTTCGGAGGAGGAGCTCTGTTTCTCCAAATGTTGTCAGCT 160  
QY 540 GluLcLcAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTptLysSerGlu 559  
DB 159 GATGAAGACGATGCGCATTTCTCGCATTCGTGATGAGAGAGGTGGAATTCGAG 100  
QY 560 LeuGlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArg 579  
DB 99 CTGAGATTGTCATGCGAAACCTTGAAGCTTCAGGCTTCAGTTAAGCTTCCTTCAGA 40  
QY 580 ValProTyrGlyPheHisGlyThrPheIle 589  
DB 39 GTTCTTATGGGTTTCACGGTACTTTATT 10

RESULT 12

AL950790  
LOCUS  
DEFINITION

Arabidopsis thaliana T-DNA flanking sequence GK-331D05-016046, genomic survey sequence.

ACCESSION AL950790.1 GI:24407412

VERSION GSS.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE 1 Strizhov,N., Li,Y., Rosso,M., Viehoever,P., Dekker,K., Saedler,H. and Weissshaar,B.  
A pipeline for automated high-throughput generation of PSTs (flanking sequence tags) from Arabidopsis thaliana T-DNA transformed lines

TITLE Unpublished

JOURNAL

REFERENCE

AUTHORS Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weissshaar,B.  
TITLE A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat) for flanking sequence tag based reverse genetics

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 667)

AUTHORS Rosso,M., Strizhov,N., Li,Y. and Weissshaar,B.  
JOURNAL Direct Submission

COMMENT Submitted (21-OCT-2002) Weissshaar B., Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany  
This sequence is recovered from the left border of the T-DNA. It indicates an insertion close to or within gene Atg30100. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at:  
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

FEATURES  
source  
1..667  
Location/Qualifiers

/organism="Arabidopsis thaliana"

/mol\_type="genomic DNA"

/strain="Columbia 0"

/db\_xref="taxon:3702"

/clone="GK-331D05-016046"

/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"

BASE COUNT 176 a 145 c 161 g 184 t 1 others  
ORIGIN

Alignment Scores:

Pred. No.: 2,57e-83 Length: 667

Score: 888.00 Matches: 164

Percent Similarity: 86.30% Conservative: 25

Best Local Similarity: 74.89% Mismatches: 30

Query Match: 28.19% Indels: 0

DB: 29 Gaps: 0

US-09-758-269-6 (1-599) x AL950790 (1-667)

QY 171 PheAspGlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerTyrAla 190

DB 7 TTCGACGGTACGGTATGTTTCACGCCCTTAAATCACTAACGGAGCGCGAGTACTCG 66

QY 191 CysArgPheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPhe 210

DB 67 TCCCGGTTTACGGAAACCGAGAGATGGTTTCAAGAGAAACCACTCGGTTCCCGATTTC 126



```

REFERENCE
AUTHORS
1 (bases 1 to 627)
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.
and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: msekic@cc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified lambda FLC-1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified pluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index\_e.html) for further
details.
TITLE
JOURNAL
COMMENT

FEATURES
source
1. .627
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL08-11-H16"
/dev_stage="rosette plants"
/lab_host="DH10B"
/clone_lib="RAFL08"
/notes="Site 1: BamHI; Site 2: SalI; subjected to
dehydration-treated (1, 2, 5, 10, 24 hr)"
BASE COUNT 162 a 179 c 128 g 158 t
ORIGIN

Alignment Scores:
Pred. No.: 637e-81 Length: 627
Score: 865.00 Matches: 168
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 27.46% Indels: 0
DB: 9 Gaps: 0

US-09-758-269-6 (1-599) x AV826228 (1-627)

Qy 1 MetAlaSerPheThrAlaThAlaAlaValSerGlyArgTrpLeuGlyGlyAsnHisThr 20
Db 124 ATGGCTTCTTTCACGGCAACGGCTGCGGTCTTCGGAGATGGCTTGGTGGCAATCATACT 183
Qy 21 GlnProProLeuSerSerSerGlnSerSerAspLeuSerTyrCysSerSerLeuProMet 40
Db 184 CAGCGCGCATATTCGTCTTCTCAAAGCTCCGACTTGAGTTATTGATGCTCTTACCTATG 243
Qy 41 AlaSerArgValThrArgGlyLeuAsnValSerSerAlaLeuHisThrProProAlaLeu 60
Db 244 GCCAGTCGTGTACACGCTAAGCTCAATGTTTCACTCGCTTCACACTCTCCAGCTCTT 303
Qy 61 HisPheProGlyGlnSerSerAsnSerProAlaLeuValValValProGlyAlaGlyGlu 80
Db 304 CATTTCCCTAAGCAATCATCAATCTCCCGCCATTTGTTAAGCCCAAGCAAGAA 363
Qy 81 SerAsnThrGlyGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAlaAla 100
Db 364 TCCACACTAACAGATGAATTTGTTCCAGAGAGCGGCGGCGAGCGTTGGACGCGCG 423
Qy 101 GluGlyPheLeuValSerHisGluGlyLeuHisProLeuProGlyThrAlaAspProSer 120
Db 424 GAGGGTTTCTTGTTCAGCCACGAGAGCTACACCGCTTCCATAAACGGCTGATCTTAGT 483
Qy 121 ValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuPro 140
Db 484 GTTCAGATCGCGGAAATTTTGTCTCCGTGTGATGAACAGCCGCTCCGCGGTATCTTCG 543

Qy 141 ValValGlyGlyLeuProAspSerIleGlyValThrValArgAsnGlyAlaLeuPro 160
Db 544 GTGGTCGGAATTCCTCCGATTTCATCAAGAGGTGATGTGGCAACGAGCTACCCA 603
Qy 161 LeuHisGluProValThrGlyHis 168
Db 604 CTCACAGCGCGGTGACAGGTCA 627

RESULT 15
CC157378 668 bp DNA linear GSS 29-APR-2003
LOCUS ig19b12.b1 WGS-Zmaysf (DH5a methyl filtered) Zea mays genomic clone
DEFINITION ig19b12, genomic survey sequence.
ACCESSION CC157378
VERSION CC157378.1 GI:30182158
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
REFERENCE
AUTHORS Rabinowicz, P., O'Shaughnessy, A.L., Ballija, V., Dedhia, N.,
Katzenburger, F., King, L., Miller, B., Muller, S., Nascimento, L.,
Zutavern, I., McCombie, W.R. and Martienssen, R.A.
Genomic shotgun sequences from Zea mays (methyl-filtered)
Unpublished
Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874
Email: mccombie@cshl.org
Plate: ig19 row: b column: 12
Seq primer: -21M13UnivFwd
Class: shotgun
High quality sequence stop: 668.
Location/Qualifiers
1. .668
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="ig19b12"
/lab_host="DH5a"
/clone_lib="WGS-Zmaysf (DH5a methyl filtered)"
/notes="Organ: immature ears; Site 1: Xba I; Site 2: Xba I;
The vector was digested with XbaI and one nucleotide was
added by fill in in the recessive 3' end. The genomic DNA
was nebulized, end repaired, adaptor ligated and size
fractionated using sephadex. The resulting fragments were
between 0.8 and 3 kb and were cloned into the vector (.x/y
reads in M13mp19, .b/g reads in pUC19). The same ligation
was transformed into DH5a."
BASE COUNT 89 a 229 c 248 g 102 t
ORIGIN

Alignment Scores:
Pred. No.: 1.02e-79 Length: 668
Score: 854.00 Matches: 163
Percent Similarity: 83.3% Conservative: 22
Best Local Similarity: 73.42% Mismatches: 34
Query Match: 27.11% Indels: 3
DB: 29 Gaps: 1

US-09-758-269-6 (1-599) x CC157378 (1-668)

Qy 188 SerTyrAlaCysArgPheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArg 207
Db 3 TCGTACGGTGTCCGGTTCCACGAGCGCGCGCTCCGCGAGGACCGCGCTGGGGGG 62
Qy 208 ProValPheProLysAlaIleGlyLeuHisGlyHisThrGlyIleAlaArgLeuMet 227

```

63	GCCTGTTTCCCAAGGCACATCGGCGAGCTCGACGGGCACTTCGGGATCCGCGCTCGCG	122
228	LeuPheTyrAlaArgAlaIleValAspProAlaHisGlyThrGlyValAla	247
123	CTGTCTACGGCGCGGGCTGTGGGGCTCGTCGACCGTCCGCGGCACGGGGTGCA	182
248	AsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAlaMetSerCluAspLeuPro	267
183	AACCGCGCCCTCGTCTACTTCAACGGCGGTCTCTGGCCATGTCGAGACACCTCGCG	242
268	TyrGlnValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAsp	287
243	TACCAAGTTCGCGTACGCGCGACGGCGACCTTCGCACCGTTCGGCGCTACGACTTCGC	302
288	GlyGlnLeuGlu-----SerThrMetIleAlaHisProLysValAspProGluSerGly	305
303	GGCGAGCTCGCGCGGTGCGGCAGCATGATCGCGACCCCAAGCTTCACCCGCGTCCGG	362
306	GluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPhe	325
363	GAGCTGTTGCGCTCAGCTACGACGTGATCAAGCGGCCCTTACCTGAGGTACTTCTACCTG	422
326	SerProAspGlyThrLysSerProAspValGluIleGlnLeuAspGlnProThrMetMet	345
423	CGGCCCGACGGCGGCAAGTCGGGGCGACGTGGAGATCCGCTGGAGCAGCCACCATGTG	482
346	HisAspPheAlaIleThrGluAsnPheValValValProAspGlnGlnValValPheLys	365
483	CACGACTTCGCGATCACGGAGCGGTTCGTGGTGGTCCGACCAACAGGTGGTGTCAAG	542
366	LeuProGluMetIleArgGlyGlySerProValValTyrAspLysAsnLysValAlaArg	385
543	CTGGCGCAGATGCTCGCGGGGGGTTCGCCCGTGTGTGTGACCGCGGCAGACGTCGCG	602
386	PheGlyIleLeuAspLysTyrAlaCluAspSer-SerAsnIleLysTrpIleAspAla	405
603	TTGCGGTGCTGCCCAAGTACGGCGCGACGCCCTCGAGATGGCGTGGGTGACGTGCC	662
405	oAsp 406	
663	GGAC 666	

Search completed: November 16, 2003, 23:12:10  
Job time : 3382 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 21:13:32 ; Search time 39 Seconds

(without alignments)  
3963.423 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150

Sequence: 1 MASPTATAVSGRWLGNGHT.....VPYGFHGTFIGADLLAKQVV 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mnc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	3150	100.0	599	Q9LR7	Q9LR7 arabidopsis
2	3143	99.8	599	Q93ZUS	Q93ZUS arabidopsis
3	2289	72.7	604	Q9M329	Q9M329 solanum tub
4	2280.5	72.4	605	Q24023	Q24023 lycopersico
5	2170.5	68.9	657	Q9M9F5	Q9M9F5 arabidopsis
6	2168.5	68.8	612	Q9FS24	Q9FS24 vigna ungui
7	2165	68.7	615	Q9M6E8	Q9M6E8 phaseolus v
8	2155.5	68.4	601	Q9LP16	Q9LP16 pisum sativ
9	2073.5	65.9	608	Q8LP15	Q8LP15 pisum sativ
10	2066.5	65.6	625	Q9AXZ4	Q9AXZ4 persea amer
11	2044.5	64.9	589	Q9C6Z1	Q9C6Z1 arabidopsis
12	1991	63.2	583	Q49505	Q49505 arabidopsis
13	1977	62.8	569	Q9AXZ3	Q9AXZ3 persea amer
14	1930	61.3	604	Q24592	Q24592 zea mays (m
15	1663.5	52.8	577	Q9LRM7	Q9LRM7 arabidopsis
16	983	31.2	574	Q8LP14	Q8LP14 pisum sativ

## ALIGNMENTS

RESULT 1

Q9LR7 PRELIMINARY; PRT; 599 AA.

AC Q9LR7, 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
DE 9-cis-epoxycarotenoid dioxygenase (Neoxanthin cleavage enzyme).  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; euclcotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Columbia;  
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Columbia;  
RA Nakamura Y.;  
RX MEDLINE=20277480; PubMed=10819329;  
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence  
RT features of the regions of 4,504,864 bp covered by sixty P1 and PAC  
RT clones.";  
RL DNA Res. 7:131-135(2000).  
[3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Col;  
RA Iuchi S., Kobayashi M., Shinozaki K.;  
RT "Characterization of neoxanthin cleavage enzyme from Arabidopsis  
RT thaliana.";  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB028617; BAB01336.1; -  
DR EMBL; AB026549; BAB70609.1; -  
DR InterPro; IPR004294; RPE65.  
DR Pfam; PF03055; RPE65; 1.  
KW Dioxygenase.

Q49675 arabidopsis  
Q94IR2 phaseolus v  
Q8LP17 pisum sativ  
Q9LY63 arabidopsis  
Q65572 arabidopsis  
Q8GR12 arabidopsis  
Q94EN8 lactuca sat  
Q9AXZ5 persea amer  
Q8YXV3 anabaena sp  
Q49695 malus domes  
Q9AA32 caulobacter  
Q9AM11 streptomyce  
Q93FA4 streptomyce  
Q06785 mycobacteri  
Q9KX46 streptomyce  
O05905 mycobacteri  
P74334 synecocyst  
Q8VXN9 crocus sativ  
Q8LNK8 oryza sativ  
Q8DMU6 synecococ  
Q8YPB4 anabaena sp  
Q53353 pseudomonas  
Q8VXP1 crocus sativ  
Q8YX26 arabidopsis  
Q8XIA2 raistonia s  
P74370 synecocyst  
Q8RQW2 pseudomonas  
Q52008 pseudomonas  
Q93VD5 oryza sativ

17 968 30.7 595 10 Q49675  
18 959.5 30.5 543 10 Q94IR2  
19 945.5 30.0 541 10 Q8LP17  
20 938 29.8 538 10 Q9LY63  
21 937 29.7 538 10 Q65572  
22 922.5 29.3 501 10 Q8GR12  
23 837.5 26.6 200 10 Q94EN8  
24 798.5 25.3 524 10 Q9AXZ5  
25 745.5 23.7 475 16 Q8YXV3  
26 745 23.7 446 10 Q49695  
27 528.5 16.8 483 16 Q9AA32  
28 527.5 16.7 456 2 Q9AM11  
29 526.5 16.7 456 2 Q93FA4  
30 465 14.8 501 16 Q06785  
31 464 14.7 503 16 Q9KX46  
32 454.5 14.4 502 16 O05905  
33 394 12.5 490 16 P74334  
34 382.5 12.1 209 10 Q8VXN9  
35 382.5 12.1 251 10 Q8LNK8  
36 358 11.4 487 16 Q8DMU6  
37 355.5 11.3 497 16 Q8YPB4  
38 352 11.2 485 2 Q53353  
39 327 10.4 218 10 Q8VXP1  
40 322.5 10.2 570 10 Q8YX26  
41 319 10.1 520 16 Q8XIA2  
42 317 10.1 480 16 P74370  
43 311.5 9.9 491 2 Q8RQW2  
44 303.5 9.6 490 2 Q52008  
45 302.5 9.6 552 10 Q93VD5

SQ SEQUENCE 599 AA; 65856 MW; 7D513F39945E0CF3 CRC64;

Query Match 100.0%; Score 3150; DB 10; Length 599;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-235;  
 Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASFTATAAAGSGRWLGNGHTQPLSSSSQSDLSYCSLPMASRVTRKLNVSALHTPPAL 60  
 DB 1 MASFTATAAAGSGRWLGNGHTQPLSSSSQSDLSYCSLPMASRVTRKLNVSALHTPPAL 60

QY 61 HFPKQSSNSPAIVVKPKAKESNTKQNLFORAAAAALDAEGLVSHEKHLPLKPTADPS 120  
 DB 61 HFPKQSSNSPAIVVKPKAKESNTKQNLFORAAAAALDAEGLVSHEKHLPLKPTADPS 120

QY 121 VOIAGNFAPVNEQVVRNLFPVVGKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAV 180  
 DB 121 VOIAGNFAPVNEQVVRNLFPVVGKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAV 180

QY 181 KFEHGSASYACRTQTNRVQERQLGRPVFPKAIAGELHGHGTIARLMLFYARAAAGIVDP 240  
 DB 181 KFEHGSASYACRTQTNRVQERQLGRPVFPKAIAGELHGHGTIARLMLFYARAAAGIVDP 240

QY 241 AHGTGVANAGLVYFNGELLANSDDLPYQVQITPNGDLKTGVRFPDGOLESTMIAPKV 300  
 DB 241 AHGTGVANAGLVYFNGELLANSDDLPYQVQITPNGDLKTGVRFPDGOLESTMIAPKV 300

QY 301 DPESGELFALSYDVVSKPYLKYFRFSPDGTKSPDVEIQDQPTMMHDFAITENFVVVPDQ 360  
 DB 301 DPESGELFALSYDVVSKPYLKYFRFSPDGTKSPDVEIQDQPTMMHDFAITENFVVVPDQ 360

QY 361 QVVPKLPKEMIRGGSPVYVDKVKARFGILDKYAEADSSNIKWIDAPCFCHLWNWEEPE 420  
 DB 361 QVVPKLPKEMIRGGSPVYVDKVKARFGILDKYAEADSSNIKWIDAPCFCHLWNWEEPE 420

QY 421 TDEVVVIGSCWTPPDSIFNESDENLKSVLSEIRLNKTGSTRRPIISNEDQOVNLEAGM 480  
 DB 421 TDEVVVIGSCWTPPDSIFNESDENLKSVLSEIRLNKTGSTRRPIISNEDQOVNLEAGM 480

QY 481 VNRNMLGRKTKFAYLALAEWPVKVSGFVKVLDLTGGEVKHLYGDNRYGGEPFLPGE 540  
 DB 481 VNRNMLGRKTKFAYLALAEWPVKVSGFVKVLDLTGGEVKHLYGDNRYGGEPFLPGE 540

QY 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSRVYPYFGHTFGIADDLAKQV 599  
 DB 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSRVYPYFGHTFGIADDLAKQV 599

RESULT 2  
 Q93ZU5 PRELIMINARY; PRT; 599 AA.

AC Q93ZU5  
 DT 01-DEC-2001 (TRENBLrel. 19, Created)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
 DE Putative 9-cis-epoxycarotenoid dioxygenase.  
 GN MOA2.4/AT3G14440.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Yanada K., Banh J., Banno F., Dale J.M., Goldsmith A.D., Lee J.M.,  
 RA Yanada K., Quach H.L., Tang C., Toriumi M., Yamamura Y., Yu G.,  
 RA Onodera C.S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,  
 RA Yu S., Bowser L., Kamiya A., Karlin-Neumann G., Kawai J., Kim C.,  
 RA Ishida J., Jones T., Kaniya A., Karlin-Neumann G., Kawai J., Kim C.,  
 RA Koesema E., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M.,  
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shimizu P.,  
 RA Southwick A., Tracy S.E., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.;  
 RT "Full length cDNA of gene MOA2.4/AT3G14440 (GI:11994214).";  
 RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY056255; AAL07104.1; -;  
 DR InterPro; IPR004294; RPE65.  
 DR Pfam; PF03055; RPE65; 1.  
 DR DIOXYGENASE  
 SQ SEQUENCE 599 AA; 65813 MW; B9007A2DC1C15506 CRC64;

Query Match 99.8%; Score 3143; DB 10; Length 599;  
 Best Local Similarity 99.8%; Pred. No. 5.7e-235;  
 Matches 598; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MASFTATAAAGSGRWLGNGHTQPLSSSSQSDLSYCSLPMASRVTRKLNVSALHTPPAL 60  
 DB 1 MASFTATAAAGSGRWLGNGHTQPLSSSSQSDLSYCSLPMASRVTRKLNVSALHTPPAL 60

QY 61 HFPKQSSNSPAIVVKPKAKESNTKQNLFORAAAAALDAEGLVSHEKHLPLKPTADPS 120  
 DB 61 HFPKQSSNSPAIVVKPKAKESNTKQNLFORAAAAALDAEGLVSHEKHLPLKPTADPS 120

QY 121 VOIAGNFAPVNEQVVRNLFPVVGKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAV 180  
 DB 121 VOIAGNFAPVNEQVVRNLFPVVGKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAV 180

QY 181 KFEHGSASYACRTQTNRVQERQLGRPVFPKAIAGELHGHGTIARLMLFYARAAAGIVDP 240  
 DB 181 KFEHGSASYACRTQTNRVQERQLGRPVFPKAIAGELHGHGTIARLMLFYARAAAGIVDP 240

QY 241 AHGTGVANAGLVYFNGELLANSDDLPYQVQITPNGDLKTGVRFPDGOLESTMIAPKV 300  
 DB 241 AHGTGVANAGLVYFNGELLANSDDLPYQVQITPNGDLKTGVRFPDGOLESTMIAPKV 300

QY 301 DPESGELFALSYDVVSKPYLKYFRFSPDGTKSPDVEIQDQPTMMHDFAITENFVVVPDQ 360  
 DB 301 DPESGELFALSYDVVSKPYLKYFRFSPDGTKSPDVEIQDQPTMMHDFAITENFVVVPDQ 360

QY 361 QVVPKLPKEMIRGGSPVYVDKVKARFGILDKYAEADSSNIKWIDAPCFCHLWNWEEPE 420  
 DB 361 QVVPKLPKEMIRGGSPVYVDKVKARFGILDKYAEADSSNIKWIDAPCFCHLWNWEEPE 420

QY 421 TDEVVVIGSCWTPPDSIFNESDENLKSVLSEIRLNKTGSTRRPIISNEDQOVNLEAGM 480  
 DB 421 TDEVVVIGSCWTPPDSIFNESDENLKSVLSEIRLNKTGSTRRPIISNEDQOVNLEAGM 480

QY 481 VNRNMLGRKTKFAYLALAEWPVKVSGFVKVLDLTGGEVKHLYGDNRYGGEPFLPGE 540  
 DB 481 VNRNMLGRKTKFAYLALAEWPVKVSGFVKVLDLTGGEVKHLYGDNRYGGEPFLPGE 540

QY 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSRVYPYFGHTFGIADDLAKQV 599  
 DB 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSRVYPYFGHTFGIADDLAKQV 599

RESULT 3  
 Q9M329 PRELIMINARY; PRT; 604 AA.

AC Q9M329  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
 DE Putative 9-cis-epoxycarotenoid dioxygenase.  
 GN NCED1.  
 OS Solanum tuberosum (Potato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamiales; Solanales; Solanaceae; Solanum.  
 OX NCBI\_TaxID=4113;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Burridge A., Taylor I.B., Thompson A.;  
 RT "Potato putative 9-cis-epoxycarotenoid dioxygenase 1 cDNA.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ276244; CAB76920.1;  
 DR InterPro; IPR004294; RPE65.  
 DR Pfam; PF03055; RPE65; 1.



RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,  
RA Shinn P., Altai H., Bei Q., Chin C., Chou J., Choi E., Conn L.,  
RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,  
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharasy N., Nguyen M.,  
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,  
RA Ecker J.R.;  
RA "Genomic sequence for Arabidopsis thaliana BAC F3F9 from chromosome  
RT I.";  
RT Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RL Ecker J.R.;  
RA Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
RA Khan S., Kim C., Altai H., Bei B., Chin C., Chou J., Choi E., Conn L.,  
RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,  
RA Lee J., Lenz C., Li J., Liu A., Liu S., Mukharasy N.,  
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
RA Theologis A., Ecker J.;  
RA Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AC013430; AF17197.1; -;  
DR InterPro; IPR004294; RPE65.  
DR Pfam; PF03055; RPE65; 1.  
SQ SEQUENCE 657 AA; 73015 MW; F41DCEB94806318 CRC64;

Query Match 68.98; Score 2170.5; DB 10; Length 657;  
Best Local Similarity 67.78; Pred. No. 1.7e-159;  
Matches 419; Conservative 80; Mismatches 87; Indels 33; Gaps 9;

Qy 1 MASPTATAVSGRWL-----GNHGTOPPLSSQSDLSYCSSLP-----MASRVT 45  
Db 51 MASPTLLPSTSTQELDTFTSTSSSRPKLQS-----LSFSTLRNKLKLVVPCVSSV 105  
Qy 46 KLVNSALHT----PALHPKSSNSPAIVKPKAKESNTKOMLFOFAAAALDAAE 101  
Db 106 KSSVSSLSQSPPTKPPSW---KLCNDVTNLI-PKTTNQPK-LNPVQRTAMVLDAVE 160  
Qy 102 GFLVSHK-KLHPLKPTADPSVQIAGNFAPVNEQVRRLNPVVGKLPDSIKGVYVRNGANP 160  
Db 161 NAMISHERRRHPKTPADPAVQIAGNFFVPEKPVVHNLPTGTVPKICQGVYVRNGANP 220  
Qy 161 LHEPVTHGHPDGDGMVHAKFEHGSASACRFTQTNRFVQEROLGRPVFPKALGELHGH 220  
Db 221 LHKPVSGHLLPDGDGMVHAKFEHGSASACRFTQTNRFVQEROLGRPVFPKALGELHGH 280  
Qy 221 TGIARLMFLFYARAAAAGIVDPAHGTGVANAGLVYFNGRLVMSDEDLFPYQVQITPNGDLKT 280  
Db 281 LGIAKMLFNTRGLFGLVDPTGGLGVANAGLVYFNGHLLAMSEDDLFPYHVKVTQTGDLT 340  
Qy 281 VGRFPDQQLSTWIAHAKVDPESGELFALSVDVYVSKPYLYKYPFSPDGTSPDVEIQLD 340  
Db 341 SGRYDFDQQLSTWIAHAKVDPESGELFALSVDVYVSKPYLYKYPFSPDGTSPDVEIQLD 400  
Qy 341 OPTMMDHFAITENFVVDQVVFVKLPEMIRGSGPVVYDKNKFARFGLDKYAEVSSNIK 400  
Db 401 OPTMMDHFAITENFVVDQVVFVKLPEMIRGSGPVVYDKNKFARFGLDKYAEVSSNIK 460  
Qy 401 WIDAPDCFCFHLWNABEPEDEVVVGSCMTPPDSIPNESDENLKSVLSEIRLNKTGE 460  
Db 461 WIEVPCFCFHLWNABEPEDEVVVGSCMTPPDSIPNESDENLKSVLSEIRLNKTGE 520  
Qy 461 STRRPIISNEDQVNLKAGVNRNMLGRKTKFAYLALAEPPKVSFGAKVDLTITGEVKKH 520  
Db 521 STRRPIISNEDQVNLKAGVNRNMLGRKTKFAYLALAEPPKVSFGAKVDLTITGEVKKH 577  
Qy 521 LYGDNRVYGGEPFLFPGEGEEDGYILCFVHDEKTKWSELOQVNAVSVLEATVKLPSRV 580  
Db 578 IYGEKYGGEFLFPGEGEEDGYIMVFVHDEKTKWSELOQVNAVSVLEATVKLPSRV 637

Qy 581 PYGFHGTGFIAGADDLAKQVV 599  
Db 538 PYGFHGTGFIKEDLSKQAL 656  
RESULT 6  
ID Q9FS24 PRELIMINARY; PRT; 612 AA.  
AC Q9FS24;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE Neoxanthin cleavage enzyme.  
GN CPRD65.  
OS Vigna unguiculata (Cowpea).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.  
OX NCBI\_taxid=3917;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20317197; PubMed=10859185;  
RA Iuchi S., Kobayashi M., Yamaguchi-Shinozaki K., Shinozaki K.;  
RT "A stress-inducible gene for 9-cis-epoxycarotenoid dioxygenase  
RT involved in abscisic acid biosynthesis under water stress in drought-  
RT tolerant cowpea";  
RL Plant Physiol. 123:553-562 (2000).  
DR EMBL; AB030293; BAB11932.1; -;  
DR InterPro; IPR004294; RPE65.  
DR Pfam; PF03055; RPE65; 1.  
SQ SEQUENCE 612 AA; 67715 MW; E3263A13293B188B CRC64;

Query Match 68.84; Score 2168.5; DB 10; Length 612;  
Best Local Similarity 68.64; Pred. No. 2.1e-159;  
Matches 423; Conservative 61; Mismatches 98; Indels 35; Gaps 8;

Qy 8 AAVSGRWLGGNHTOPPSSQSDLSYCSSLPMASTRKLN-VSSALHTPPALHFPKQ- 65  
Db 3 SSASNTWENATLPSPPPKDLPTS-SPTNLLPL-RKTSSTNTTCSLOT---LHFPKQY 56  
Qy 66 ----SSNSPAIVKPKAKESNT-----KOMLFOFAAAALDAAE 100  
Db 57 QSTSTSTATTTPTPKTTITTTTPPRETNPLSDTNQPLPKWFLQKAAATALLDV 116  
Qy 101 EGFVSHKELHPLKPTADPSVQIAGNFAPVNEQVRRLNPVVGKLPDSIKGVYVRNGANP 160  
Db 117 ETALVSHKELHPLKPTADPSVQIAGNFAPVNEQVRRLNPVVGKLPDSIKGVYVRNGANP 176  
Qy 161 LHEPVTHGHPDGDGMVHAKFEHGSASACRFTQTNRFVQEROLGRPVFPKALGELHGH 220  
Db 177 LYEPVAGHHFPDGDGMVHAKFEHGSASACRFTQTNRFVQEROLGRPVFPKALGELHGH 236  
Qy 221 TGIARLMFLFYARAAAAGIVDPAHGTGVANAGLVYFNGRLVMSDEDLFPYQVQITPNGDLKT 280  
Db 237 SGIAKMLFNTRGLFGLVDPTGGLGVANAGLVYFNGHLLAMSEDDLFPYHVKVTQTGDLT 296  
Qy 281 VGRFPDQQLSTWIAHAKVDPESGELFALSVDVYVSKPYLYKYPFSPDGTSPDVEIQLD 340  
Db 297 VGRYDFDQQLSTWIAHAKVDPESGELFALSVDVYVSKPYLYKYPFSPDGTSPDVEIQLD 356  
Qy 341 OPTMMDHFAITENFVVDQVVFVKLPEMIRGSGPVVYDKNKFARFGLDKYAEVSSNIK 400  
Db 357 EPTMMDHFAITENFVVDQVVFVKLPEMIRGSGPVVYDKNKFARFGLDKYAEVSSNIK 416  
Qy 401 WIDAPDCFCFHLWNABEPEDEVVVGSCMTPPDSIPNESDENLKSVLSEIRLNKTGE 460  
Db 417 WIDAPDCFCFHLWNABEPEDEVVVGSCMTPPDSIPNESDENLKSVLSEIRLNKTGE 476  
Qy 461 STRRPIISNEDQVNLKAGVNRNMLGRKTKFAYLALAEPPKVSFGAKVDLTITGEVKKH 520  
Db 477 STRRPIISNEDQVNLKAGVNRNMLGRKTKFAYLALAEPPKVSFGAKVDLTITGEVKKH 535  
Qy 521 LYGDNRVYGGEPFLFPGEGEEDGYILCFVHDEKTKWSELOQVNAVSVLEATVKLPSRV 580

```
Db 536 MYGEKFGGEPLFLP-NGQKDDGYILAFVHDEKWKSELOIVNAQNKLKLEASIKLPSRV 594
Qy 581 PYGPHGTFTICADDLAKQ 597
Db 595 PYGPHGTFTIHSKDLRKQ 611

RESULT 7
Q9M6E8
ID Q9M6E8 PRELIMINARY; PRT; 615 AA.
AC Q9M6E8;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE 9-cis-epoxycarotenoid dioxygenase.
GN NCED1.
OS Phaseolus vulgaris (Kidney bean) (French bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
OX NCBI_TaxID=3885;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Top Crop;
RX MEDLINE=20079657; PubMed=10611388;
RA Qin X., Zeavaart J.A.;
RT "The 9-cis-epoxycarotenoid cleavage reaction is the key regulatory
step of abscisic acid biosynthesis in water-stressed bean.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:15354-15361(1999).
DR EMBL; AF190462; AAF26356.1; -.
DR InterPro; IPR004294; RPE65.
DR Pfam; PF03055; RPE65; 1.
KW Dioxygenase.
SQ SEQUENCE 615 AA; 68075 MW; 0CC10F862D7DE130 CRC64;

Query Match 68.7%; Score 2165; DB 10; Length 615;
Best Local Similarity 69.2%; Pred. No. 4e-159;
Matches 426; Conservative 59; Mismatches 95; Indels 36; Gaps 9;

Qy 11 SGRWLGNGHTQPLSSQSSDLSCYSS-----LPMASRVTRKLN-VSSALHTPPALHPFK 64
Db 6 SNTWI--NTTLSSCSSPKDLAS:SSSPTTLLPKKRSSSNTNITCSLQI---LHYPK 60
Qy 65 Q-----SNSPAIVVVKPAKE-----SNTKQ-----NMLFORAAALDAAE 101
Db 61 QYQPTSTTTTPTIKPTTTTTTTPHRETKPLSDTKQFPQKWNFLQKAAATGLDMVE 120
Qy 102 GFVVSHEKHLPLPKTADPSVQIAGNAPVNEQPVRRNLPVVGKLPDSIKGVYVRNGANPL 161
Db 121 TALVSHESKHLPLPKTADPKVQIAGNAPVPEHAADQALPVGKI PKCIDGVYVRNGANPL 180
Qy 162 HEPVTGHHFFDGMVHAVKFEHGSASACRTQTNRFRVQERQOLGRPVFPKALGELHGT 221
Db 181 YEPVAGHHFFDGMVHAVKFTNGAASACRETTQRLAQEKSLGRPVFPKALGELHGS 240
Qy 222 GTARLMLFVARAAGVDPAGHGTGVANAGLVFENGELAWSDDLPYQVQITPENGDLKT 281
Db 241 GTARLLTFARSLFQVLVDGSHGVANAGLVFNNHLLAWSDDLPYHVRITNSGLTIV 300
Qy 282 GREDFDQGLESTMIAPKVDPSGELFALSVDVSKPYLYKFRFSPDGTGKSPDVEIQLDQ 341
Db 301 GRYDFNGQLNSTMIAPKLDPVNGDLHALSYDVQKPYLYKFRFSGADGVKSPDVEIPLKE 360
Qy 342 PTMHHDFATENFVVPDQGVFKLPDMIRGSGSPVYDKNKVARGLDKYAEADSNIKW 401
Db 361 PTMHHDFATENFVVPDQGVFKLTEMITGSGSPVYDKNKTSRFGILDKNKADNMRW 420
Qy 402 IDAPDFCFHLNWAWEPEPTEVVWIGSMTTPDSIFNESDENLKSIVLSEIRLNKTGBS 461
Db 421 IDAPECFCHLNWAWEPEPTEIVWIGSMTTADSI FNECDSELSKIVLSEIRLNKTGS 480
Qy 462 TRRPILISNDDQVNLBAGMVRNMLGRKTKFAYLALAEWPVKVSGPAKVDLTFSGEVK 521
```

```
Db 481 TRRPILIS-DAEQVNLBAGMVRNMLGRKTKFAYLALAEWPVKVSGPAKVDLTFSGEVK 539
Qy 522 YGDNRYGGEPLFLPGEGGEDEGYILCFVHDEKWKSELOIVNAVLEVEATVKLPSRVP 581
Db 540 YGEEKGGEPLFLP-NGEEGDGYILAFVHDEKWKSELOIVNAQNKLKLEASIKLPSRVP 598
Qy 582 YGPHGTFTICADDLAKQ 597
Db 599 YGPHGTFTIHSKDLRKQ 614

RESULT 8
Q8LP16
ID Q8LP16 PRELIMINARY; PRT; 601 AA.
AC Q8LP16;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Nine-cis-epoxycarotenoid dioxygenase2.
GN NCED2.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX NCBI_TaxID=3886;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Alaska; TISSUE=Axillary bud;
RA Nakako A., Mori H.;
RT "Isolation and characterization of a cDNA encoding a nine-cis-
epoxycarotenoid dioxygenase2(pSNCE2) from Pisum sativum.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB080192; BAC10550.1; -.
DR InterPro; IPR004294; RPE65.
DR Pfam; PF03055; RPE65; 1.
KW Dioxygenase.
SQ SEQUENCE 601 AA; 67287 MW; F4EDC257867CF195 CRC64;

Query Match 68.4%; Score 2155.5; DB 10; Length 601;
Best Local Similarity 68.7%; Pred. No. 2.1e-158;
Matches 422; Conservative 61; Mismatches 100; Indels 31; Gaps 8;

Qy 1 MASFTATAVSGRWLGNHTQPLSSQSSDLSCYSSLPASRVTRKLNVSALHTPPAL 60
Db 1 MATFTASPP--SNTW-----INTKSSRTSSSILLNKKRSTKNTISCSLQI--TL 47
Qy 61 HFPK-----QSSNSPAIVVVKPAK-----ESNTKQM-----NLFORAAAAALDAEGL 104
Db 48 PFPKPYQPSNTTITTLITRETKNLPSTNPKLHKQEKQKLLQAAAATLDPVETTL 107
Qy 105 VSHEKHLPLPKTADPSVQIAGNAPVNEQPVRRNLPVVGKLPDSIKGVYVRNGANPLHEP 164
Db 108 IKOESKHLPLPKTADPSVQIAGNAPVPEHPTQNLPTITGKLPKIDGVYLRNGANPLHEP 167
Qy 165 VTGHHFFDGMVHAVKFEHGSASACRTQTNRFRVQERQOLGRPVFPKALGELHGTGIA 224
Db 168 VAGHHFFDGMVHAVKFTNGSVSYSCRTTETHRLAQEKALGRPVFPKALGELHGSIA 227
Qy 225 RLMLFYARAAAGVDPAGHGTGVANAGLVYENGELAWSDDLPYQVQITPENGDLKTGVR 284
Db 228 RLMLFYARSLCGLVDGTHGMGVANAGLVYENNLAWSDDLPYHVRVTENGDLTTCVY 287
Qy 285 DFDSQGLESTMIAPKVDPSGELFALSVDVSKPYLYKFRFSPDGTGKSPDVEIQLODPTM 344
Db 288 DFNDQKSTMIAPKVDPVKNLYALSVDVQKPYLYKFRFDSNGVSKSPDVEIPLAEP 347
Qy 345 MHDFAITENFVVPDQGVFKLPDMIRGSGSPVYDKNKVARFGLDKYAEADSNIKWIDA 404
Db 348 MHDFAITENFVVPDQGVFKLPDMIRGSGSPVYDKNKVARFGLDKYAEADSNIKWIDA 407
Qy 405 PDCEFCFLNWAWEPEPTEVVWIGSMTTPDSIFNESDENLKSIVLSEIRLNKTGSTRR 464
```

```

408 PECFCFLMNAWEEPENDEVVVIGSWTTPADSIIFNECDLSKSVLSVLEIRLNLTGKSTR 467
465 PITSNEDQVNLKAGVNRNMLGRKTKFAYLALAEPPKVSQFAKVDLTGGVKKHLYCD 524
468 AII-QSEHNNLSEAGVNRNMLGRKTKFAYLALAEPPKVSQFAKVDLTGGVKKHLYGE 526
525 NRYGGSPFLFPGGGE-EDRGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPGRVPYG 583
527 NRGGEPLFLPNEDSENEDDGYILTFFVHDEKWKSEIQIVNAVNLKLEACIPLPSRVPYG 586
584 FHGTFCADDLAKQ 597
587 FHGTFTHSKLEKQ 600

RESULT 9
Q8LP15 PRELIMINARY; PRT; 608 AA.
ID Q8LP15
AC Q8LP15
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Nine-cis-epoxycarotenoid dioxygenase3.
GN NCED3
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Alaska; TISSUE=Axillary bud;
RA Nakako A., Mori H.;
RT "Isolation and characterization of a cDNA encoding a nine-cis-
RT epoxycarotenoid dioxygenase3 (PNCED3) from Pisum sativum."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB080193; BAC10551.1;
DR InterPro; IPR004294; RPE65.
DR Pfam; PF03055; RPE65; 1.
KW Dioxygenase.
SQ SEQUENCE 608 AA; 67978 MW; 73E80913AAA26C4D CRC64;

Query Match 65.9%; Score 2075.5; DB 10; Length 608;
Best Local Similarity 68.3%; Pred. No. 3.4e-152;
Matches 403; Conservative 64; Mismatches 98; Indels 25; Gaps 5;

QY 24 LSSQSSDLSYCSLPAASVTRKLVNVSALHTTPALHFFKQSSNSP-----AIVV 74
DB 28 ISSSSSTSFNOKSTLKLKKPKRKLQCAVHSFSLDYPKQSYKQPLINKNDTETIIH 87
QY 75 KPAKESNTKQNLQFQRAAAALDAAGFLVSHKHLPLPKTADPSVQIAGNFAPVNEQP 134
DB 88 KPK----ESSQWNPLOKAAALANWFESALLSRELVPLPKTSDPRIQIAGNFAPVPEQ 143
QY 135 VRNLPPVVKLPDSIKGVYVNGANPLHEPTVGHFFDGGGMHAKFEHGSASYACRFT 194
DB 144 VVHSLFVTKGKPRCVNGVYVNGANPMFEPVSGHLLFDGGGMHAKVNTNDGVASYACRFT 203
QY 195 QTNRRVQRQLGRPVFPKAIAGELHGHGTGIALMLFYARAAAGIYDPAHGTGVANAGLVYF 254
DB 204 ETERLVQERELGRAMFPKAIAGELHGHGTGIALMLFYARSLCGIYDHRGSGVANAGLVYF 263
QY 255 NGRLLAMSEDDLPYQVQITPNGDLKTVGRFPDQGLESTMTAHKVPDPESGELFALSYD 314
DB 264 DGKLLAMSEDDLPYELQTPYQGLTKVGRYSFLDQLHSTMTAHKPIDIPISGELFALSYE- 322
QY 315 VSKPYLVFRSPDGTGSPDVEIQDQPTMHDFAITENFVVDPQVVKLPKPMIRGGS 374
DB 323 VARPYLVFRSPDGKSPDVEIRLSVTMTLHDAITENYVYIPDHQVVKLEEMIKGS 382
QY 375 PVVYDNKVARPGLDKVAEASSNKTWIDAPDCFCFLHNAWESPETDEVVVIGSCMTPR 434
DB 383 PVIFDGAKSRFGLPKYAKDASSIIVWDSPTDFCFLHNAWESPETAIVVIGSCMTPP 442

```

```

435 DSIFNESDENLKSVLSEIRLNLTGSTRPPIISNEDQVNLKAGVNRNMLGRKTKFAY 494
443 DSIFNESDENLKSVLSEIRLNLTGSTRSIV-----PQNLBAGVNRNMLGRKTKFAY 498
495 LALAEPPKVSQFAKVDLTGGVKKHLYGDNRYGGSPFLFPGS-----GGDESGYIL 547
499 LAVAEPPKVSQFAKVDLESVEGKHSYGDGRFGGEPFLPFRSGRCEVENEEDGGYIM 558
548 CFVHDEKTKWSELQIVNAVSLVEATVKLPGRVPYGFHGTFCADDLAKQ 597
559 ALVHDERCKSELQIVNAVSLVEATVKLPGRVPYGFHGTFCVAKDLTFQ 608

RESULT 10
Q9AXZ4 PRELIMINARY; PRT; 625 AA.
ID Q9AXZ4
AC Q9AXZ4;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE 9-cis-epoxycarotenoid dioxygenase.
GN NCED3.
OS Persea americana (Avocado).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Laurales; Lauraceae; Persea.
OX NCBI_TaxID=3435;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Lula;
RA Chernys J., Zeevaert J.A.D.;
RT "Abasic acid in avocado fruit."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF224671; AAK00623.1;
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF03055; RPE65; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Dioxygenase.
SQ SEQUENCE 625 AA; 69724 MW; A56F33042D5F2CB6 CRC64;

```

```

Query Match 65.6%; Score 2066.5; DB 10; Length 625;
Best Local Similarity 64.9%; Pred. No. 1.8e-151;
Matches 392; Conservative 74; Mismatches 93; Indels 45; Gaps 6;

QY 38 LPMASRVTRKLVNVSALHTTP-----ALHFF-----KQSS 67
DB 22 LPISKNLSRPKNFIMLKXNTPLIQCSHSPSSSAAVLHLPKQPTKSKPSIKKGSST 81
QY 68 NSPAIVVVKAKESNTKO-----MNLFORAAALDAAGFLVSH--EKLHPLPKT 116
DB 82 LTPSIEKPNPGSHQVKTQSGPNRVGNWNIIFORTAAFALDAIEKLIARVLERHPLPKT 141
QY 117 ADPSVQIAGNFAPVNEQPVRRNLPVVGKLPDSIKGVYVNGANPLHEPTVGHFFDGGGM 176
DB 142 ADPEVQIAGNFAPVAEHVQHGIPVAGRIPCLDGVYVNGANPLFEBIAGHHFFDGGGM 201
QY 177 VNAVFEHGSASYACRFTQTNRFQEROLGRVPKAIAGELHGHGTGIALMLFYARAAAG 236
DB 202 IHAVFRNGSASYSRYTETRLRQEROLSRPIPKAIAGELHGHGTGIALMLFYTRGLFG 261
QY 237 IYDPAHGTGVANAGLVYVNGRLAMSEDDLPYQVQITPNGDLKTVGRFPDQGLESTMTA 296
DB 262 LVNADEGVANAGLVYVNRLLAMSEDDLPVHVEITPSGDLKTVGRHDFDQLASSMIA 321
QY 297 HPKVPDPESGELFALSYDVVSKPYLVKYPFRSPDGTGSPDVEIQDQPTMHDFAITENFV 356
DB 322 HPKLDPESGELFSLSYDVARKPYLVKYPHAPDGWKSPPDVEIPDRFTMIHDAITENFV 381
QY 357 VPDQVVKLPKPMIRGSPVVYDNKVARFGILDKVAEASSNKTWIDAPDCFCFLHNAW 416
DB 382 IPDQVVKLEEMIRGSPVYDNKVARFGILDKVAEASSNKTWIDAPDCFCFLHNAW 441

```







```

QY 75 -KPKAKESNTKONNLFORAAAAALDA-ABGFLVS-HEKHLPLKPTADPSVOIAGNAPVN 131
Db 72 AAKRAEGGKKLNLFORAAAAALDAFEBGFVANLERPHGLPSTADPAVQIAGNAPFVG 131
QY 132 EQPVARNLFWGKLPDSIKGVVVRNGANPLHPEVPTGHHFFDGDGMVHAKVEHGS-AYA 190
Db 132 ERPPVHELFPVSGRIIPFPIDGVYARNGANPCFPDPVAGHLLFDGDGMVHALRIRNGAESYA 191
QY 191 CRFTQTNRFVQBRQGRPVFPKAIAGELHGHGTGIALMLFYARAAAAGIVDPAHGTGVANAG 250
Db 192 CRFTETARLQBRRAIGRPVFPKAIAGELHGHSGIARLALFYARAAAGLVDPDSAGTGVANAG 251
QY 251 LVYFNGRLAMGDEDDLPLYOVQVITPNDGLKTVGRFDFDQGLESTMIAPKVPDPESGELPAL 310
Db 252 LVYFNGRLAMGDEDDLPLYHVRVADDDGLTVGRYDFDQGLGCAMTAHPKLPATGELHAL 311
QY 311 SYDVSVKPLKYLFRSPDGTSPDVEIQLDQPTMHDFAITENFVVVDQVQVFKLPPEMI 370
Db 312 SYDVSVKPLKYLFRSPDGTSPDVEIQLDQPTMHDFAITENFVVVDQVQVFKLPPEMI 371
QY 371 RGSFVVYDKNVKVARGIIDKVAEDSSNIKWIDAPDCFCFHLNNAWEEBETDEVVVIGSC 430
Db 372 RGSFVVYDKNVKVARGIIDKVAEDSSNIKWIDAPDCFCFHLNNAWEEBETDEVVVIGSC 431
QY 431 MTPPDSIFNESDENLKSVLSEIRLNLKTGSTRPPIISNEDQOVNLEAGMVRNMLGRKT 490
Db 432 MTPADSIENESDERLESVLTIRLDARTGRSTRRAVLP-PSQENLENVGMVRNMLGRES 490
QY 491 KEAYLALAEPMKVGKFAKVDLTTCGEVKHLYGDNRYGGEPLFLPGEGE-----EDGEG 545
Db 491 RYAYLVAEPWPKESGFAKEDJSTGELTKFYEYGEGRFGGEPFCFVPMDPAAAHPRGEDDGY 550
QY 546 ILCFVHDEXTWKSQELQVNNVSLVEATVKLPSRPVYGFHGTGFIAGDDLAKQV 598
Db 551 VLTFFVHERAGTSELVAVNNAADIRLEATVQLPSRPVYGFHGTGFIAGDLEAQ 602

```

## RESULT 15

```

Q9LRW7 PRELIMINARY; PRT; 577 AA.
AC Q9LRW7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 9-cis-epoxycarotenoid dioxygenase.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones."
RL DNA Res. 7:131-135(2000).
DR ENBL; AB028621; BAB01363.1; -.
DR InterPro; IPR004294; RPE85.
DR Pfam; PFO3055; RPE65; 1.
KW Dioxygenase.
SQ SEQUENCE 577 AA; 63821 MW; 515E0A1B321B4FE1 CRC64;

```

Query Match 52.8%; Score 1663.5; DB 10; Length 577; Best Local Similarity 55.4%; Pred. NO. 2.8e-120; Matches 323; Conservative 100; Mismatches 137; Indels 23; Gaps 9;

Search completed: November 14, 2003, 21:17:39  
Job time : 42 secs

```

QY 24 LSSQSQSDLSYCSSLPMASRVTRKLNVS-ALHTPPALHPKOSNSPAIVVKPKAKESN 82
Db 9 LLFTKTSRSHLLPQPKANISRIILINPKIPTLPDLTSPVESP-----VKLKETYPN 62
QY 83 TKOWNLFORAAAAALDAAB-GFLVSEKHLPLPKTADPSVOIAGNAPVNEQOVRRNLVP 141
Db 63 ---LNLLQKLAAMTMDKIESSIVIPNEQNRPLPKPTDPAVOLSGENAFVNECEFPQNGLEV 119
QY 142 VGKLPDSIKGVVVRNGANPLHPEVPTGHHFFDGDGMVHAKV--FEHGSASYACRFTQTNRF 199
Db 120 VGOITPSCLKGVYIRNGANPMFPPLAGHLLFDGSGMTHAVSIGFDN-QVSYSCRYTTNRL 178
QY 200 VQBRQGRPVFPKAIAGELHGHGTGIALMLFYARAAAAGIVDPAHGTGVANAGLVYFNGRL 259
Db 179 VQETALGRSVFPKAIAGELHGHSGIARLALFTARAGIGLVDTGRGMGVANAGVVFNGRLL 238
QY 260 AMSEDDLPLYOVQVITPNDGLKTVGRFDFDQGLESTMIAPKVPDPESGELFALSVDVVSXPY 319
Db 239 AMSEDDLPLYOVKIDGQGLLETIGRFGFDQDIDSVAHPKVDATTGLTLSTNVLKKPH 298
QY 320 LKYFRSPDGTSPDVEIQLDQPTMHDFAITENFVVVDQVQVFKLPPEMIRGSGFVVVD 379
Db 299 LRVLKFNCTCKKTRDVEITLPEFTMHDFAITENFVVPDQKVFKLSEIRMGSPVIYV 358
QY 380 KKVAREGIIDKVAEDSSNIKWIDAPDCFCFHLNNAWEEBETDE----VVVIGSCMTPPD 435
Db 359 KEKMARFGLSKQDLTGSDINWVDVDFDCFCFHLNNAWEE-RTTEGDPVIVVIGSCMSPD 417
QY 436 SIFNESDENLKSVLSEIRLNLKTGSTRPPIISNEDQOVNLEAGMVRNMLGRKTKEYL 495
Db 418 TIFSEGEPTFRVSELSEIRLNMRTKESNRKVIVTG---VNLEAGHINRSYVGRKSQFVI 473
QY 496 ALAEPWPVKSGFAKVDLTTCGEVKHLYGDNRYGGEPLFLPGEGEDEGYILCFVHDEKT 555
Db 474 AIADPWPKCSGIKVDIQNGTVSEFNVGPSRFGGEPFCFVPEGEDEKGYVMGFVRDEK 533
QY 556 WKSELOIVNNAVSLVEATVKLPSRPVYGFHGTGFIAGDDLAKQV 598
Db 534 DSEFVVVDATDMKQVAAVRLPERVPYGFHGTGFIAGDLEAQ 576

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 21:12:47 ; Search time 17 Seconds  
(without alignments)  
1657.000 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150

Sequence: 1 MASTATNAVSGRWLGNHT.....VPYGFHGTTCADDLAKQVV 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	120	3.8	616	1 NXF1_COTJA	P58797 coturnix co
2	114	3.6	870	1 YCS0_YEAST	P25623 saccharomyc
3	113	3.6	466	1 DNAA_PROMI	P22837 proteus mir
4	107.5	3.4	658	1 PAX1_SCHPO	P50527 schizosacch
5	106	3.4	414	1 STG6_HUMAN	O94864 homo sapien
6	106	3.4	3562	1 PGCV_CHICK	Q90953 gallus gall
7	105	3.3	783	1 CAP5_MOUSE	P55284 mus musculu
8	105	3.3	2504	1 PAS_HUMAN	P43327 homo sapien
9	104.5	3.3	1079	1 APCE_FREDI	P16566 fremyella d
10	103	3.3	775	1 SYFB_AQUAE	O67620 aquifex aeo
11	102	3.2	513	1 M113_MOUSE	Q8Bgt6 mus musculu
12	102	3.2	2493	1 CYAA_USTWA	P49606 ustilago ma
13	101.5	3.2	520	1 GLGS_BRANA	Q9m462 brassica na
14	101.5	3.2	562	1 SYR_XANCP	O8p455 xanthomonas
15	101.5	3.2	565	1 SCH1_SCHPO	O14335 schizosacch
16	101	3.2	512	1 GLGT_VICFA	P52417 vicia faba
17	100	3.2	3770	1 ACVS_EMENI	P27472 emericeila
18	99.5	3.2	383	1 PHYC_BACSU	O31097 bacillus su
19	99.5	3.2	1256	1 ATL_STRAU	P52081 staphylococ
20	99.5	3.2	4705	1 FAT2_DROME	Q9vw71 drosophila
21	98	3.1	741	1 MAS2_WYCTU	O50596 mycobacteri
22	98	3.1	1633	1 YP74_CAEEL	O93221 caenorhabdi
23	97.5	3.1	518	1 CP74_ARATH	Q98242 arabidopsis
24	97.5	3.1	691	1 VP33_YEAST	P20795 saccharomyc
25	97.5	3.1	730	1 EF2_METAC	Q8trc3 methanosarc
26	97	3.1	385	1 OXDC_BACSU	O34714 bacillus su
27	97	3.1	636	1 KDGP_SCHPO	O14019 schizosacch
28	96.5	3.1	426	1 UCR2_SCHPO	P18761 schizosacch
29	96.5	3.1	503	1 DLTA_PACSU	P39581 bacillus su
30	96.5	3.1	862	1 SMP2_YEAST	P32567 saccharomyc
31	96.5	3.1	3587	1 TYCE_BREPA	O30408 b tyrocidin
32	96	3.0	462	1 DNAA_YERPE	Q829u7 yersinia be
33	95.5	3.0	865	1 NRFA_PENUR	Q92269 penicillium

## ALIGNMENTS

### RESULT 1

ID	NXF1_COTJA	STANDARD;	PRT;	616 AA.
AC	P58797	2003 (Rel. 41, Created)		
DT	28-FEB-2003	(Rel. 41, Last sequence update)		
DT	28-FEB-2003	(Rel. 41, Last annotation update)		
DE	Nuclear RNA export factor 1 (Tip associating protein) (Tip-associated protein) (mRNA export factor TAP).			
DE	NXF1			
GN	NXF1			
OS	Coturnix coturnix japonica (Japanese quail).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Coturnix			
OX	NCBI_TaxID=93934;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND MUTAGENESIS OF GLU-246.			
RX	MEDLINE=21257892; PubMed=11358864;			
RA	Coturn G.A., Wiegand H.L., Kang Y., Ho D.N., Georgiadis M.M.,			
RA	Cullen B.R.;			
RT	"Using viral species specificity to define a critical protein/RNA interaction surface."			
RL	Genes Dev. 15:1194-1205 (2001).			
CC	-1- FUNCTION: Involved in the nuclear export of mRNA from the nucleus to the cytoplasm.			
CC	-1- SUBCELLULAR LOCATION: Nuclear.			
CC	-1- DOMAIN: The leucine-rich repeats and the NTF2-domain are essential for the export of mRNA from the nucleus (By similarity).			
CC	-1- MISCELLANEOUS: Changing a single residue, Glu-246 to Arg, the residue found in human NXF1, fully rescues both constitutive transport element (CTE) function and CTE binding.			
CC	-1- SIMILARITY: BELONGS TO THE NXF FAMILY.			
CC	-1- SIMILARITY: Contains 4 leucine-rich (LRR) repeats.			
CC	-1- SIMILARITY: Contains 1 NTF2 domain.			
CC	-1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.			
CC	-1- SIMILARITY: Contains 1 UBA-like domain.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	EMBL; AF343749; AAX58910.1; -			
DR	InterPro; IPR001611; LRR.			
DR	InterPro; IPR003603; LRRcap.			
DR	InterPro; IPR002075; NTF2.			
DR	InterPro; IPR005637; TAP_C.			
DR	Pfam; PF00560; LRR; 2.			
DR	Pfam; PF02136; NTF2; 1.			
DR	Pfam; PF03943; TAP_C; 1.			
DR	SMART; SM00446; LRRcap; 1.			
DR	PROSITE; PS50177; NTF2_DOMAIN; 1.			
KW	Transport; mRNA transport; Nuclear protein; Repeat;			

Q9jhy8 rattus norv  
Q9jxz2 neisseria m  
Q9civ5 mus musculu  
Q97hd0 clostridium  
P29555 drosophila  
P16310 rattus norv  
P87074 schizosacch  
P55118 pasteurella  
P96177 weissella h  
P55801 mycoplasma  
Q94967 homo sapien  
Q9cew0 lactococcus





```
DR EMBL; U22371; AAC49125.1; ..
DR EMBL; AL034433; CAA22347.1; ..
DR EMBL; L41552; AAB52609.1; ..
DR FJ; S60170; S60170.
DR FJ; T39500; T39500.
DR HSP; Q00534; IBI8.
DR GeneB_Spomb; SPSC1604.14c; ..
DR InterPro; IPR000095; PAKBox/RhoBndng.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR002290; Ser Thr pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Piam; P000786; PDB; 1.
DR Piam; P000659; Pkinase; 1.
DR PRINTS; P00109; TYRKINASE.
DR PtoDom; PD00001; Prot kinase; 1.
DR SMART; SM00285; PBD; 1.
DR SMART; SM00220; S_TK; 1.
DR PROSITE; PS0108; CRIB; 1.
DR PROSITE; PS0107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS0011; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS0108; PROTEIN KINASE_ST; 1.
DR Transferase; Serine/threonine-protein kinase; ATP-binding;
KW Phosphorylation.
FT DOMAIN 147 160 CRIB.
FT DOMAIN 85 88 POLY-SER.
FT DOMAIN 234 237 POLY-SER.
FT DOMAIN 246 249 POLY-SER.
FT DOMAIN 263 267 POLY-SER.
FT DOMAIN 386 637 PROTEIN KINASE.
FT NP_BIND 392 400 ATP (BY SIMILARITY).
FT BINDING 415 415 ATP (BY SIMILARITY).
FT ACT_SITE 505 505 BY SIMILARITY.
FT CONFLICT 492 495 GLOH -> LYSD (IN REF. 1).
FT CONFLICT 537 537 R -> P (IN REF. 1).
FT SEQUENCE 658 AA; 73258 MW; 69D72E5C925021E5 CRC64;

Query Match 3.4%; Score 107.5; DB 1; Length 658;
Best Local Similarity 20.6%; Pred. No. 3;
Matches 88; Conservative 61; Mismatches 157; Indels 121; Gaps 20;

Qy 18 NHTQPLSSQSSDL-----SYCSLLPWSRVTRKLVSSALHPPALHFPKQSS 67
Db 237 SHLQPTATSSSSSLYPSRPAPTPASSSSPLSSQTVKTTTNSASRQSPVLVSKSTD 296

Qy 68 N-----SPAIVVFKAKESNTKQ-----NLFQRAAAALDAAEGFLVSHKHLPLPKTA 117
Db 297 NIIRSHSPVLLTPQTLSTSEKTRPNNSTPYQRAETSK-----PKAV 341

Qy 118 DPSVQIAGNPAFVNEQVRRNLVVGKLPDSIKGVYVRNGANPLHEPVTHGHFFDGDGMV 177
Db 342 ATPQKVEAPSAFRLQKRAPRO-----QSNDSAVLAKLOSICNPKNPTLLYRNF----- 389

Qy 178 HAVKFEHGSAS--YACRFTQNRVQERQLGRPVFPK-----AIGEL-----HGHTGIARLM 227
Db 390 --VKGQASGDVYSARQVGNLSVAIKMNIQPKKEFVNEILLVMKSHHKNIVFI 447

Qy 228 --LYARAAGIVDPAHGT-----VANAGLVYFNGRLLMSGDDLPVQVQITPNG-----D 277
Db 448 DTFYKSELMMVMYMEGSGSLTEVVTNNTL--SEGQAAICKETLEGQLHENGIVHRD 505

Qy 278 LKT-----VGRFDFDGOLES-----TWIAHP-----KVDPEES 304
Db 506 IKSDNILLSLQDGIKLTDFGCAQIDSNMKTETTMVGTFFYMAPEVTVRKEYGFKVDVMS 565

Qy 305 GELFALSYDVVSKPYLK-----YFRFPDGTQSPDVEIQDQTM-----HDAITENF 354
Db 566 LGIMAIEVVEGPPYLNENPLRALYLIATIGTP-----KISRPELLSSVFFHDP-LGSKL 618

Qy 355 VVVPDQ 361
Db 619 TVNPKQR 625
```

```
RESULT 5
ST66_HUMAN STANDARD; PRT; 414 AA.
AC O94864; O94864;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE STAGA complex 65 Gamma subunit (Adenocarcinoma antigen ART1).
GN KIAA0764.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE=20441578; PubMed=10987294;
RA Nishizaka S., Gomi S., Harada K., Oizumi K., Itoh K., Shichijo S.;
RT "A new tumor-rejection antigen recognized by cytotoxic T lymphocytes
infiltrating into a lung adenocarcinoma.";
RL Cancer Res. 60:4830-4837(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
RL DNA Res. 5:277-286(1998).
RN [3]
RP SEQUENCE OF 6-19, AND IDENTIFICATION AS PART OF THE STAGA COMPLEX.
RX MEDLINE=21448975; PubMed=11564863;
RA Martnez E., Palhan V.B., Tjernberg A., Lyman E.S., Gamper A.M.,
RA Kundu T.K., Chait B.T., Roeder R.G.;
RT "Human STAGA complex is a chromatin-acetylating transcription
coactivator that interacts with pre-mRNA splicing and DNA
damage-binding factors in vivo.";
RL Mol. Cell. Biol. 21:6782-6795(2001).
CC -! SUBUNIT: Component of the STAGA transcription coactivator-HAT
complex.
CC -! SUBCELLULAR LOCATION: Nuclear.
CC -! ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC Name=2;
CC IsoId=O94864-1; Sequence=Displayed;
CC IsoId=O94864-2; Sequence=VSP_003974;
CC Note=No experimental confirmation available;
CC -! TISSUE SPECIFICITY: Expressed at high levels in adenocarcinomas
and gliomas and low in esophageal cancers and malignant
hematological disease. Also expressed at high level in the thymus,
low in peripheral blood mononuclear cells, and lowest in the
stomach, small intestine, and skeletal muscle.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF197954; AAG28523.1; ..
CC EMBL; AF224759; AAG47636.1; ..
CC EMBL; AB018307; BAA34484.1; ..
CC GO; GO:0005634; C:nucleus; NAS.
CC InterPro; IPR006565; BTP.
CC SMART; SM00576; BTP; 1.
CC Transcription regulation; Nuclear protein; Antigen;
CC Alternative splicing.
CC VARSPLIC 1 4 MNLQ -> ML (in isoform 2).
CC FT FTId=VSP_003974.
```





[illegible]

FT CARBOHYD 440 440 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 57 69 KDO -> GKIK (IN REF. 2).  
 SQ SEQUENCE 783 AA; 87847 MW; 240AF2D663BCE71C CRC64;

Query Match  
 Best Local Similarity 17.7%; Pred. No. 5.9;  
 Matches 128; Conservative 101; Mismatches 242; Indels 254; Gaps 32;

QY 60 LHFFQSQSNPAIVVVKPAKSNYQKMLFQRAAAA-----ALDAEGFLVSEKHLHP----- 112  
 DB 52 MHIDEKQESLPHYKYDQSNVNRQNAKYVLOQEGFAGKIFGVDANTGNVLAERLDREKVS 111  
 QY 113 -----LPKTADPSVQ-----IAGNPAPVNEQPVRRNLPPVVGKLPDSIKGVY 153  
 DB 112 EYFELTALVDKNTKNLEQPSFTVKVHDINDNWFVSHQVFNASVPMSAIGTSVIRVT 171  
 QY 154 VRNGANP-----LHEPVTQHFF--DGDGNVHAVKFEHGSASYACRFQTQVRFVBER 203  
 DB 172 AYVADDDPTVAGHATVLYQIVKNEYFSIDNSGLI-----FTKIKNLDREK 216  
 QY 204 QLGRPVFEPK--AIGELHGHGTGIALMLFYARAAAAGIVDPAGHTGVANAGLVYFNGRLLA 260  
 DB 217 QAEYKIVETQALG-LREGSATVMI-----RL-- 245  
 QY 261 MSEDLLPYQVQITPNDGLTKVGRFPDQGLESTIAHP-----KYDPSGELFALSVDV 315  
 DB 246 --ED-----INDNFPVFTQYTFSPV-EDIRVGKPLGFLTVDPDPQNRMTKYSIM 295  
 QY 316 SXPLKLYFRFSPD-----GTSPDVE-----IQLDQPTMHDF----- 348  
 DB 296 QEYRDTFTETDPKNEGIIKPTSLDYEVIQOYTFYIEATDPIRVEYLSSTSGKKA 355  
 QY 349 AITENFVVVPDQVW-----FKLPEMIRG---GSPVVYDKNKVAR-FGILDKYAEDSS- 397  
 DB 356 MVTINVLVDPEPVPQRFHFHFKLPENQKKPLIGTVAKDPKQAKRSIGYSIRKTSRQ 415  
 QY 398 -----NIKMTDAPDCFFHLNWAESPETDE-----VVVIGSCWT 432  
 DB 416 FFEITKQNIYNEKELDRETYAWNL--TVANELDSRNPVKESIVQVYIEVDENDN 473  
 QY 433 PPSDI-----FNESDENLKSVLSEIRLMKTGSETRPIISNEQQ 473  
 DB 474 PFPEAQYEPKVCENAAQGLVVQISATDKDVVNPKEFKALK-NEDSNFTLINHNDT 532  
 QY 474 VNL--BAGMVNRNMLGRKTKFAYLALA---EPMPKVSQFAKVDLTGTEVKKHLYGDNRYG 528  
 DB 533 ANITVKYGFQNR---EHAKEHYLPVLISDNGVPSLTGTS--TLTVGVCKNEQGEFTFC 586  
 QY 529 GEPLFLPGEGBEGEYILCFV-----HDE 553  
 DB 587 EEMAAQAGVSIQALVAIFLCILITVITLLIILRRIRKQAHAKSKSALEIHEQLVYDE 646  
 QY 554 K-----TWKSELQIVNAV-----SLEVEATVKLPSPVPYGFHGTFIGADD 593  
 DB 647 EGGGEMDTTSDYDVSVLNSVAGSGTKPLURSTMDARPAVYTVQVQKPPPLAFGLHG--GPPE 703  
 QY 594 LAKOV 598  
 DB 704 MATMI 708

STANDARD; PRT; 2504 AA.

RESULT 8  
 FAS\_HUMAN  
 ID\_FAS\_HUMAN  
 AC P9327; 1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Fatty acid synthase (EC 2.3.1.85) [includes: EC 2.3.1.38; EC 2.3.1.41; EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14]  
 DE EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14  
 GN FASN OR FAS.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID:9606;  
 RN [1].  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=96004605; PubMed=7567999;  
 RA Jayakumar A., Tai M.-H., Huang W.-Y., Al-Feel W., Hsu M.,  
 RA Abu-Elheiga L., Chirala S.S., Wakil S.J.,  
 RT "Human fatty acid synthase: properties and molecular cloning";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:8695-8699(1995).  
 RN [2].  
 RP SEQUENCE OF 753-758 AND 1285-1297.  
 RX MEDLINE=94294385; PubMed=8022791;  
 RA Khatjida F.P., Jenner K., Wood F.D., Hennigar R.A., Jacobs L.B.,  
 RA Dick J.D., Pasternack G.R.;  
 RT "Fatty acid synthesis: a potential selective target for  
 RT antineoplastic therapy";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:6379-6383(1994).  
 CC -1- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF  
 CC LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.  
 CC THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN  
 CC ACYL CARRIER PROTEIN.  
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a  
 CC long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).  
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +  
 CC acetyl-[acyl-carrier protein].  
 CC -1- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +  
 CC malonyl-[acyl-carrier protein].  
 CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-  
 CC carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +  
 CC [acyl-carrier protein].  
 CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +  
 CC NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.  
 CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]  
 CC = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.  
 CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-  
 CC 2,3-dehydroacyl-[acyl-carrier protein] + NADPH.  
 CC -1- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-  
 CC carrier protein] + oleate.  
 CC -1- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.  
 CC -1- TISSUE SPECIFICITY: PROMINENT EXPRESSION IN BRAIN, LUNG, AND  
 CC LIVER.  
 CC -1- MISCELLANEOUS: THE RELATIVELY LOW BETA-KETOACYL SYNTHASE ACTIVITY  
 CC MAY BE ATTRIBUTABLE TO THE LOW 4'-PHOSPHOPANTHETHEINE CONTENT OF  
 CC THE PROTEIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 CC EMBL; U26644; AAC50259.1; -;  
 CC PIR; A57788; A57788.  
 CC Genew; HGNC:3594; FASN.  
 CC MIM; 600212; -;  
 CC InterPro; IPR001227; Ac transferase.  
 CC InterPro; IPR002085; Adh\_zn\_family.  
 CC InterPro; IPR007994; Ketosyl-synt.  
 CC InterPro; IPR006163; PP\_bind.  
 CC InterPro; IPR006162; Ppantrn attach.  
 CC InterPro; IPR000051; SAM\_bind.  
 CC InterPro; IPR001031; Thioesterase.  
 CC Pfam; PF00698; Acyl\_transf; 1.  
 CC Pfam; PF00107; Adh\_zinc\_N; 1.  
 CC Pfam; PF00109; ketoacyl-synt; 1.  
 CC Pfam; PF02801; ketoacyl-synt\_C; 1.  
 CC Pfam; PF00550; pp-binding; 1.  
 CC Pfam; PF00975; Thioesterase; 1.

RESULT 9	APCE_FREDI	STANDARD;	PRT; 1079 AA.
ID	APCE_FREDI		
AC	P16566;		
DT	01-AUG-1990 (Rel. 15, Created)		
DT	01-AUG-1990 (Rel. 15, Last sequence update)		
DT	01-DEC-1992 (Rel. 24, Last annotation update)		
DE	Phycobilisome 120 kDa linker polypeptide, core (L-CM 92) (Core-membrane linker protein).		
GN	APCE		
OS	Fremyella diplosiphon (Calothrix PCC 7601).		
OC	Bacteria; Cyanobacteria; Nostocales; Microchaetaceae; Fremyella.		
OX	NCBI_TaxID=1197;		
RN	{1}		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE=90192765; Pubmed=2107546;		
RA	Houmard J., Capuano V., Colomano M.V., Coursin T., de Marsac N.;		
RT	"Molecular characterization of the terminal energy acceptor of the cyanobacterial phycobilisomes.";		
PL	Proc. Natl. Acad. Sci. U.S.A. 87:2152-2156(1990).		

Query Match	3.38;	Score 104.5;	DB 1;	Length 1079;		
Best Local Similarity	21.3%;	Pred. No. 10;				
Matches	71;	Conservative	80;	Indels 135; Gaps 21;		
QY	62	FPKQSSN----	SPATVVPKAKESNTKQMNLFQRAAAAALDAAGFLVSHKLP-----	L 113		
		:	:			
		:	:			
DB	467	FPKETRNPNTSPA----	PFSKDTTRILIN-----	OGFGINSQVSNFGARGEF 509		
		:	:			
		:	:			
QY	114	PKTADPSV----	QIAG-----	NPAVPNEQPVRRNLPVVGKLPDSIKGVYVRNG 157		
		:	:			
		:	:			
DB	510	PGSLGPKVFRLDQLPGTIGKKAAGASIKTSSTQAV-----	IKAAYLQ-----	554		
		:	:			
		:	:			
QY	158	ANPLHEPVTGHHFDGDM--	VHAKVPEHGSAS-----	YACRFT 194		
		:	:			
		:	:			
DB	555	-----	VFGRDVYEGORLQVQEKLENGQLSVREFIRALAKSDVFRKTYNTSLVKVCAI	607		
		:	:			
		:	:			
QY	195	QTNRFVQERQGRPVFPKRAIGELGHGHTGIARLMLFYARAAAGIVDPAHGTGVANAGLVYF		254		
		:	:			
		:	:			
DB	608	E--YIHRLLGRPTYGRQ--	EINKYFDIAAKQGFYA-----	VVD-----	AIINS--	VEY 650
		:	:			
		:	:			
QY	255	NGRLLAMSEDLPLQVQITPNP----	DLKTVG--	RFDFDQGLESTMIHAKPKVDPSGELF		308
		:	:			
		:	:			
DB	651	SE--AFGEDTVPYERYLTPSGVALRQLR--	VGSIREDVGG-----	KVQKQETPLF		696
		:	:			
		:	:			
QY	309	ALSYDWSKPKLYKFRFSPDGTGSPDVBQLDQ		341		

Matches	114;	Conservative	67;	Mismatches	159;	Indels	212;	Gaps	31
QY	123	TAGNAPVNEOPVRNLPVVKGLPDSIKGVYVRNGANPLHEPVTGHHFFD							172
DB	46	VFGKVVVEKHEPTKKLAVV							86
QY	173	GDGMVHVKFEHGSASYACRTQTNRVQERQLGRPVFPKPAIGELHGHGTIGARLMLFYAR							232
DB	87	GDGVI--VALPNAKVGNMC--VTEREF							121
QY	233	AAAGIVDPAHGTGVANAGLVYFNGRLAMSEDDLP							277
DB	122	QELGLEEKSEG							167
QY	278	LKTVGRFDFDQLESTMIHAKVPD							331
DB	168	MUSVGVARD--LSAIFLPKKPEEPTYEETGEFFIEIEDECKY							216
QY	332	SPDVEIQLDQP							378
DB	217	ISGVEIK-ESPLYIKKRLWCGIKSINNVDITNVV							262
QY	379	DNKVARGILDKYAEDSNKWDADPCFCFHLWNWEPEETDEVVVGSCMTTPPSIF							438
DB	263	DLISKV-EGGIIVRSKXGKEKITLDG							309
QY	439	N-----ES--DENLKSVLSEI							478
DB	310	GVIGLEGISIKENTDILLESAYNPPFRVKASKKLGIOLESSYR							362
QY	479	GMVNRMLGRKTKFA-YIALAEPPKVGSKFAKVDLTGTEVKKHLYGDNRYGGEPLFLP							535
DB	363	RVDRADQYAVVILKHAGKV							404
QY	536	GEGBEDEGYILCFVHDEKTKWSELIQVNAVSLVEATVTKLPSRVPYGFHGF							588
DB	405	KYIRVAGESYKNEE							448
QY	589	IGAD-DLAKQVV							599
DB	449	IQRDVDLTBEIM							460
RESULT 11									
M113 MOUSE									
ID	M113 MOUSE	STANDARD;	PRT;	513 AA.					
AC	Q8BGT6;								
DT	15-SEP-2003 (Rel. 42, Created)								
DT	15-SEP-2003 (Rel. 42, Last sequence update)								
DT	15-SEP-2003 (Rel. 42, Last annotation update)								
DE	Molecule interacting with Rab13 (MIRAB13) (Fragment)								
GN	MIRAB13.								
OS	Mus musculus (Mouse)								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.								
OX	NCBI_TaxID=10090;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=CS7BL/6J; TISSUE=Testis;								
RX	MDLINE=22354683; PubMed=12466851;								
RA	Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,								
RA	Nikaido I., Osato N., Saito R., Suzuki H., Yamakawa I., Kiyosawa H.,								
RA	Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojoberi T.,								
RA	Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,								
RA	Schram L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,								
RA	Blake J.A., Bratt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,								
RA	Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,								
RA	Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,								
RA	Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,								
RA	Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,								
RA	Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,								
RA	Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,								
RA	Nagashima T., Numata K., Okido T., Pavan W., Pertea G., Pesole G.,								

RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sardinia A., Schneider C., Sempere C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan A., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.,  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs";  
RL Nature 420:563-573 (2002);  
CC -!- SUBUNIT: Binds to Rab13 (By similarity).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AK077217; BAC36689.1; -  
CC EMBL; AK076763; BAC36472.1; -  
CC  
CC FT NON TER 1 75 POLY-GLU.  
CC FT DOMAIN 67 75 POLY-GLU.  
CC FT DOMAIN 459 462 POLY-GLU.  
CC  
CC SQ SEQUENCE 513 AA; 55721 MW; 70CD36AC49CB061 CRC64;  
  
Query Match 3.2%; Score 102; DB 1; Length 513;  
Best Local Similarity 22.8%; Pred. No. 5.3;  
Matches 66; Conservative 35; Mismatches 82; Indels 106; Gaps 17;  
  
QY 18 NHTQPLSSSSDLSYSSLPMSRVTRKLNVSALHT-----PPALHPKQSSNSP 70  
DB 149 SHSEPP-----SATPSALSVESLSSESSHTANAEPLFPV- -PK-SSSDP 193  
  
QY 71 AI-----VVKPKAESNTKQMLFORAAALDA-AEG----- 102  
DB 194 AVHVPGTGTSNSVTPSANGSLSSGELQPSGEQMLQARTKGSAGTSTKPFSGATPT 253  
  
QY 103 FLVSHKELHLPK--TADPSVQIAGNFA--PVNEQPVRLPVVVKLPDSIKGVYVRNG 157  
DB 254 PELLAGDR-NPAPPVGSASPOLQTKSKCKENPFNRKPSASPTVRK-----ATKG 303  
  
QY 158 ANPLHEPVTGHHFFDGDGWHAVKPEHGSAYACRFTQTNRVQERQLGRVPFPAIGEL 217  
DB 304 AKVPVPPAGHGF-----PLIK-----RKVQADQVPEEDI-----XGEM 338  
  
QY 218 HGHGTIARLMFLYARAAGIVDPANGTGVANAGLVFNGRLLANSEDDL 266  
DB 339 ---DNTERQL--DALEHSGVLEEKLRGANEG-----SEDDM 371  
  
RESULT 12  
CYAA\_USTWA STANDARD; PRT; 2493 AA.  
AC P49606;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Adenylate cyclase (EC 4.6.1.1) (ATP pyrophosphate-lyase) (Adenylyl  
DE cyclase).  
GN UAC1 OR REM1.  
OS Ustilago maydis (Smut fungus).  
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;  
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.  
OX NCBI\_TaxID=5270;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=518;  
EX MEDLINE=95087882; PubMed=7995519;  
RA Gold S., Duncan G., Barrett K., Kronstad J.W.;  
RT "CAMP regulates morphogenesis in the fungal pathogen Ustilago  
RT maydis";  
RL Genes Dev. 8:2895-2816 (1994).  
CC -!- FUNCTION: Plays essential roles in regulation of cellular  
CC metabolism by catalyzing the synthesis of a second messenger,  
CC CAMP.  
CC  
CC -!- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.  
CC -!- COFACTOR: Binds 1 magnesium ion per subunit (By similarity).  
CC -!- SIMILARITY: Belongs to the adenylyl cyclase class-3 family.  
CC -!- SIMILARITY: Contains 21 leucine-rich (LRR) repeats.  
CC -!- SIMILARITY: Contains 1 PP2C-like domain.  
CC -!- SIMILARITY: Contains 1 Ras-associating domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L33918; AA557469.1; -  
CC PIR; A55481; A55481. G cyclase.  
CC InterPro; IPR001854; G\_cyclase.  
CC InterPro; IPR001511; LRR.  
CC InterPro; IPR003591; LRR typ.  
CC InterPro; IPR001932; PP2C-like.  
CC InterPro; IPR00159; RA domain.  
CC Pfam; PF00211; Guanylate\_cyc; 1.  
CC Pfam; PF00560; LRR; 16.  
CC Pfam; PF00481; PP2C; 1.  
CC SMART; SM00044; CYCC; 1.  
CC SMART; SM00369; LRR\_TYP; 2.  
CC SMART; SM00332; PP2C; 1.  
CC SMART; SM00314; RA; 1.  
CC PROSITE; PS01225; GUANYLATE\_CYCLASES\_2; 1.  
CC PROSITE; PS02000; RA; 1.  
KW Lyase; Repeat; Leucine-rich repeat; CAMP biosynthesis; Metal-binding;  
KW Magnesium.  
FT DOMAIN 970 1072 RAS-ASSOCIATING.  
FT REPEAT 1084 1106 LRR 1.  
FT REPEAT 1108 1131 LRR 2.  
FT REPEAT 1132 1155 LRR 3.  
FT REPEAT 1157 1178 LRR 4.  
FT REPEAT 1179 1201 LRR 5.  
FT REPEAT 1202 1225 LRR 6.  
FT REPEAT 1227 1248 LRR 7.  
FT REPEAT 1249 1271 LRR 8.  
FT REPEAT 1273 1295 LRR 9.  
FT REPEAT 1314 1337 LRR 10.  
FT REPEAT 1338 1360 LRR 11.  
FT REPEAT 1361 1384 LRR 12.  
FT REPEAT 1386 1407 LRR 13.  
FT REPEAT 1408 1430 LRR 14.  
FT REPEAT 1432 1455 LRR 15.  
FT REPEAT 1509 1529 LRR 16.  
FT REPEAT 1533 1556 LRR 17.  
FT REPEAT 1557 1580 LRR 18.  
FT REPEAT 1581 1604 LRR 19.  
FT REPEAT 1606 1628 LRR 20.  
FT REPEAT 1633 1656 LRR 21.  
FT DOMAIN 1722 2001 PP2C-LIKE.  
FT DOMAIN 2002 2493 CATALYTIC.  
FT DOMAIN 759 763 POLY-ASP.  
FT DOMAIN 888 897 POLY-ALA.  
FT METAL 2063 2063 MAGNESIUM (BY SIMILARITY).  
FT METAL 2105 2105 MAGNESIUM (BY SIMILARITY).  
SQ SEQUENCE 2493 AA; 271979 MW; 106A872C3C1C5B58 CRC64;

```
Query Match 3.2%; Score 102; DB 1; Length 2493;
Best Local Similarity 21.7%; Pred. No. 54;
Matches 85; Conservative 44; Mismatches 136; Indels 126; Gaps 17;

QY 4 FTATAVSGRWLGGNHT-----QPLSSSQSDLSYCSLPMASRVTRKLNVSALHTP 57
DB 801 FKXTASLAGSRGTDSDVPLTALPLPGSKVDFAANKVDVLOQ-TNNLAQSALVQO- 858

QY 58 PALHFPKQSNPAIVVKPKAKESNTKQNLFORAAAALDAEGLVSH-----EKLHPL 113
DB 859 ---QSQSNHQPSPNVRTSRGAGAHM--PASAGASAAAAGKGLHHRPSKRMVNR 913

QY 114 PXTAD-----PSVQIAGNFAFVNPQVVRNLPVVGKLPDSIKVYVNGANPLHEPTG 167
DB 914 PNTAGSVGATRSPTTLGSLTSAEDD-----TSINGIRDDG-HPLKRSATA 959

QY 168 -----HFFDGDGMHVAKPEHGSASYACRTQTNRVQERQLGRPVFPKAIGE 216
DB 960 NTNNTATGTLPRNH-----IRVYKTDGTFATLSCLSVSTANEVQ----- 998

QY 217 LHGHTCIARLMFYARAAGIVDPAGHTGVANAGLVYFNGRLLAMSEDDLPYQVQ----- 271
DB 999 -----TLAKRSUTTESAAVLF-----VRDKG-----SERFLGIS--DKPSQLQRRKLI 1041

QY 272 ----ITPNDGLKTVGRDFDQGLESTMIAPKVDPSGELFALSVDWSPKYLKYRFRSPD 328
DB 1042 QAGYTENDGLEDGMRDLS-----YLLRFVFRPD 1070

QY 329 GYKSPDVE-IQDQPTMHDFAITENFVVVP 358
DB 1071 SVPTFDSIGHSHTFQHLDLHRSNLEMPV 1101

RESULT 13
GLGS_BRANA STANDARD; PRT; 520 AA.
ID GLGS_BRANA AC Q9M462;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glucose-1-phosphate adenyltransferase small subunit, chloroplast
DE precursor (EC 2.7.7.27) (ADP-glucose synthase) (ADP-glucose
DE pyrophosphorylase) (AGPase B) (Alpha-D-glucose-1-phosphate adenyl
DE transferase).
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3706;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Drakkar; TISSUE=Seed;
RA Zawodny S., Martini N.;
RT "Isolation and analysis of a cDNA clone encoding the small subunit of
RT ADP-glucose pyrophosphorylase in the plastids of seeds and leaves of
RT Oilseed rape (Brassica napus).";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: This protein plays a role in synthesis of starch. It
CC catalyzes the synthesis of the activated glycosyl donor, ADP-
CC glucose from GLC-1-P and ATP.
CC -!- CATALYTIC ACTIVITY: ATP + alpha-D-glucose 1-phosphate =
CC diphosphate + ADP-glucose.
CC -!- ENZYME REGULATION: ACTIVATED BY 3'PHOSPHOGLYCERATE, INHIBITED
CC BY ORTHOPHOSPHATE. ALLOSTERIC REGULATION.
CC -!- PATHWAY: Starch biosynthesis.
CC -!- SUBUNIT: Heterotrimer.
CC -!- SUBCELLULAR LOCATION: Chloroplast (By similarity).
CC -!- TISSUE SPECIFICITY: Leaves.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL AND PLANTS GLUCOSE-1-
CC PHOSPHATE ADENYLYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ271162; CAB89863.1; -
DR InterPro; IPR005836; ADP_Glu_pyrop.
DR Pfam; PF00483; NTP transferase; 1.
DR PROSITE; PS00808; ADP_GLC_PYROPHOSPH_1; 1.
DR PROSITE; PS00809; ADP_GLC_PYROPHOSPH_2; 1.
DR PROSITE; PS00810; ADP_GLC_PYROPHOSPH_3; 1.
KW Glycogen biosynthesis; Transferase; Nucleotidyltransferase;
KW Multigene family; Starch biosynthesis; Allosteric enzyme;
KW Chloroplast; Transit peptide.
FT TRANSIT 1 71
FT CHAIN 72 520
FT SEQUENCE 520 AA; 57044 MW; D0EAF9706F3B6A7 CRC64;
FT SMALL SUBUNIT.
FT DOEAF9706F3B6A7 CRC64;
SQ SEQUENCE 520 AA; 57044 MW; D0EAF9706F3B6A7 CRC64;
Query Match 3.2%; Score 101.5; DB 1; Length 520;
Best Local Similarity 27.0%; Pred. No. 5.9;
Matches 43; Conservative 27; Mismatches 66; Indels 23; Gaps 7;

QY 1 MASFTATAVSGRWLGNHTQPLSSSQSDLSYCSLPMASRVTRKLNVSALHTPPAL 60
DB 1 MATMAAGSLKVPSSSSNHTRLSSSQSKTLSPSSS---SLTGKLNPTQIIIS--- 53

QY 61 HFPKQSNPAIVVKPKA-KESNTKQNLFORAAAALDAEGLV---SHEKLHPL-PK 115
DB 54 NLPRGNERRTPSIVSPKAVSDSONQTCLDPDASRVL---GIILGGAGTGLYPLTKK 109

QY 116 TADPSVQIAGNFAFVNPQVVRNLPVVGKLPDSIKGVYV 154
DB 110 RAKPAVELGANYRLI-----DIPVSNCLNSNISKIYV 141

RESULT 14
SVR_XANCP
ID SVR_XANCP STANDARD; PRT; 562 AA.
AC Q8F455;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Arginyl-L-tyrosine synthetase (EC 6.1.1.19) (Arginine--trna ligase) (ArgRS).
GN ARG5 OR KCC3861
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=2022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., Bl-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Gregio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.C.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Secubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
```





```
FT DOMAIN 164 167 POLY-GLN.
FT DOMAIN 375 378 POLY-SER.
SQ SEQUENCE 565 AA; 59713 MW; FE6CCCF7E54A8CDB CRC64;

Query Match
Best Local Similarity 3.2%; Score 101.5; DB 1; Length 565;
Matches 74; Conservative 38; Mismatches 147; Indels 81; Gaps 15;

QY 6 ATAAVSGRLGGNHTQPPLLSSOSSDLS--YCSSLPMASRVTKLNVSALHTP---PAL 60
Db 147 ATAAVMSY--PHYSASVQQQQTFFVNGQPHNLPAQAPATIYGIPDALHTTQNGTTI 204
QY 61 HFPKQSSNSPAIVVKPAKESNTKQMNLFQRAAAAALDAAGFLVSHEKHLPLPKTADPS 120
Db 205 HVTGTTPPGAVSORSEPDRLSSNMENOLLASAAANQLDAA-----PRIT-PT 250
QY 121 VQIAGNFAPVNEOP-VRRNLPVVVKLPDSTIKGYVVRNGANPLH-EPVTGHHFFDGDGMVH 178
Db 251 KSSGVNLMPLUSNAPSPPKQNVVGSIPSS-----SNTSPNHLASVPNRLTSNSSTGS 303
QY 179 AVKFEHGSASYACRFTQTRRFVQEROLGRPVFP---KAIGELHGHTGIALMLFYARAAA 235
Db 304 FTKVTNGS-----SNSLYSNSMQTPYLPKSKNSSTSLHSMYGVGT----- 344
QY 236 GIYDPAHGTCVANAGLVYFNGRLLAMSEDDLPYQVOITPNGDLKTVGRFDFDQOLESTMI 295
Db 345 -----TAYAPQSURYAH-----YNYLPYRPSVNG-----FDDSSSDF 380
QY 296 AHPKVDPESGELFALSVDVVSFKYKFRFSPDGTKSPDV 335
Db 381 AHFRYQRRSRPVSPCS-TAPSSPTFTSRGSP-----TPDV 415
```

Search completed: November 14, 2003, 21:16:46  
Job time : 19 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 21:14:17 ; Search time 21 Seconds

(without alignments)  
2743.097 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150

Sequence: 1 MASFTATAVSGRWLGNGH.....VPYGFHGTFIGADDLAKQVV 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76.\*

1: piri.\*

2: piri2.\*

3: piri3.\*

4: piri4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2289	72.7	604	T51936	probable 9-cis-epo
2	2280.5	72.4	605	T07123	nine-cis-epoxycaro
3	2170.5	68.9	657	E96812	protein F3P9.10 [i
4	2044.5	64.9	589	A86425	probable 9-cis-epo
5	1991	63.2	583	T04531	nine-cis-epoxycaro
6	1930	61.3	604	T04351	viviparus-14 prot
7	968	30.7	595	T04438	hypothetical prote
8	938	29.8	538	T49193	neoxanthin cleavag
9	937	29.7	538	T51734	neoxanthin cleavag
10	745.5	23.7	475	AG1944	hypothetical prote
11	745	23.7	446	T17019	hypothetical prote
12	528.5	16.8	483	E87345	conserved hypothet
13	466	14.8	501	T470534	hypothetical prote
14	454.5	14.4	502	A70582	hypothetical prote
15	394	12.5	490	S76169	hypothetical prote
16	355.5	11.3	497	AE2341	lignostilbene-alph
17	352	11.2	484	TJ0595	lignostilbene-alph
18	317	10.1	480	S76206	hypothetical prote
19	303.5	9.6	489	JC4324	lignostilbene-alph
20	253	8.0	616	T10688	hypothetical prote
21	244	7.7	618	C84885	hypothetical prote
22	242	7.7	472	AG2417	hypothetical prote
23	238.5	7.6	556	F88115	conserved hypothet
24	231	7.3	483	D87290	retinal pigment m
25	215	6.8	533	A47143	hypothetical prote
26	126	4.0	790	T34293	hypothetical prote
27	116.5	3.7	1832	T31113	mucin-like glycopr
28	114	3.6	870	T74291	hypothetical prote
29	113	3.6	466	IOEBV	replication initia

30	110.5	3.5	658	2	S60170	protein kinase Pak
31	108	3.4	437	2	G70019	conserved hypothet
32	107.5	3.4	658	2	T39500	serine/threonine-s
33	107	3.4	1461	2	T13157	mitotic checkpoint
34	107	3.4	26926	1	I38344	titin, cardiac mus
35	106.5	3.4	1299	2	T47182	hypothetical prote
36	106	3.4	1679	2	T15968	hypothetical prote
37	106	3.4	3562	2	A47171	chondroitin sulfat
38	105.5	3.3	2591	2	T30288	pristinamycin I sy
39	105	3.3	2504	1	A57788	MYC-related DNA bi
40	104.5	3.3	623	2	T52293	enoyl-facyl-carrie
41	104.5	3.3	1080	2	A35088	phycobilisome link
42	103.5	3.3	394	2	AG3616	hypothetical prote
43	103	3.3	775	2	B70449	phenylalanine-trna
44	103	3.3	4845	2	T31067	BIR repeat contain
45	102.5	3.3	531	2	AH3378	cell wall degradat

ALIGNMENTS

RESULT 1

T51936

Probable 9-cis-epoxycarotenoid dioxygenase [imported] - potato

C:Species: Solanum tuberosum (potato)

C:Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 20-Oct-2000

C:Accession: T51936

R:Burridge, A.; Taylor, I.B.; Thompson, A.

submitted to the EMBL Data Library, March 2000

A:Description: Potato putative 9-cis-epoxycarotenoid dioxygenase 1 cDNA.

A:Reference number: 225874

A:Accession: T51936

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-604 <BUR>

A:Cross-references: EMBL:AJ276244; PIDN:CAB76920.1

C:Genetics:

A:Gene: nced1

Query Match 72.7%; Score 2289; DB 2; Length 604;  
Best Local Similarity 71.3%; Pred. No. 1.4e-165;  
Matches 432; Conservative 66; Mismatches 90; Indels 18; Gaps 5;

QY	5	TATAVSGRWLGNGHNTOPPLSSSSQSDLSYCS---	SLPMASRVTRKLNVSALHTPPALH	61
DB	3	TTTSHATNTWI-----RKPLSMPSSKEFGFASNSISLLKQHNRSQSLNINSSLAQPPILH	57	
QY	62	FPKSSSN-----SPAIVVKPKAKESNT-----KQMLFORAAAAALDAEGFLVSEKHLPL	113	
DB	58	FPKSSNYQTPTKSTISHPKQENNSSSSISKWNLVQKAAAMALDAYEGALTKHELEHPL	117	
QY	114	PKTADPSVQIAGNEAPVNEQPVRRNLPVVGKLPDSIKGVYVRNGANPLHPVPTGHHFFDG	173	
DB	118	PKTADPRVQISGNFAPVNEQPVRRNLPVVGKLPDSIKGVYVRNGANPLFEFTAGRHFFDG	177	
QY	174	DMWHAVKFEGHSYACRFTQTNRFVQEROLGRPVFPKALGELHGHGTGIALMLFYARA	233	
DB	178	DMWHAVKFEGHSYACRFTQTNRFVQEROLGRPVFPKALGELHGHGTGIALMLFYARG	237	
QY	234	AAGIVDPAGHGTGANAGLVYNGRLLAMSEDDLPVQVQITPNGDLTKTVGRDFDGOLEST	293	
DB	238	LFGLIDHGRGTGANAGLVYNNRLLAMSEDDLPVQVQITPNGDLTKTVGRDFDGOLEST	297	
QY	294	MAIHPKVDPESGELFALSVDVSKPYLKYFRFSPDGTSPDVEIQDQFTMMHDFAITEN	353	
DB	298	MAIHPKVDPESGELFALSVDVSKPYLKYFRFSPDGTSPDVEIQDQFTMMHDFAITEN	357	
QY	354	FVVVDDQVQVFKLPMIRGGSPVVDKKNVAFGLDKYAEKSSNKNWIDAPDCFCFHLW	413	
DB	358	FVVIIDQVQVFKVSSMIRGGSPVVDKKNVAFGLDKYAEKSSNKNWIDAPDCFCFHLW	417	
QY	414	NAWEEPEDEVVIGSCMTTPDPSIFNESDENLKVSLSEIRLNKLTGSTRPFIISNEDQQ	473	

Db 418 NAWEEPETDEIVVIGSCMTTPDIFNECEDEGLKSVLSEIRLNLKTGKSTRKAIENPDEQ 477  
QY 474 VNLEAGMVRNMLGRKTKFAYLALAEFPWPKVSGFAKVDLTITGEVKKHLYGDNRYGGEPLF 533  
Db 478 VNLEAGMVRNMLGRKTKFAYLALAEFPWPKVSGFAKVDLTITGEVKKHLYGDNRYGGEPLF 537  
QY 534 LPGE--GGDEDEGYILCFVHDEKTKSELQIVNAVSLVEATVKLPSRVPYGFHGTFFGA 591  
Db 538 LPRDPNSKEEDDGYILAFVHDEKTKSELQIVNAVSLVEATVKLPSRVPYGFHGTFFGA 597  
QY 592 DDLAKQ 597  
Db 598 NDLANQ 603

RESULT 2  
T07123  
nine-cis-epoxycarotenoid dioxygenase - tomato  
N:Alternate names: probable neoxanthin cleavage enzyme  
C:Species: Lycopersicon esculentum (tomato)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 20-Jun-2000  
C:Accession: T07123  
R:Burbidge, A.  
submitted to the EMBL Data Library, January 1998  
A:Reference number: Z15934  
A:Accession: T07123  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-605 <BUR>  
A:Cross-references: EMBL:297215; PIDN:CAB10168.1

Query Match 72.4%; Score 2280.5; DB 2; Length 605;  
Best Local Similarity 71.7%; Pred. No. 6.2e-165;  
Matches 435; Conservative 63; Mismatches 90; Indels 19; Gaps 6;

QY 5 TATAAAGGRLGNGHPTPPSSSSQSSDLSSYCS---SLPMASRVTRKLNVSALHTPPALH 61  
Db 3 TTTSHANTWTF-----TKLMSPKSKGFGFASNSISLLKQNRQSLNINSLSQAPILH 57  
QY 62 FPKQSSN--SPA--IVVKPKAKESN-----TKQNLFRAAAAALDAEGLVSHKELHP 112  
Db 58 FPKQSSNVQTPKNNTISHPKQENNNSSSSSTSKNVLQKAAAMALDAVESALTKHELEHP 117  
QY 113 LPKTADPSVQIAGFAPVNEOPVRNLPVVKLPDSIKGVVVRNGANPLPEPVTGHFFD 172  
Db 118 LPKTADPRVQISGNFAPVNPVQCSLPVTKIPKCVGVVVRNGANPLPEPVTAGHFFD 177  
QY 173 GDGMVHAVKFEHGSASACRFTQNRFFVQERQLGRPVFPKPAIGELHGHGTGIARLMLFYAR 232  
Db 178 GDGMVHAVQPKNGSASACRFTETELVQERKALGRPVFPKPAIGELHGHGTGIARLMLFYAR 237  
QY 233 AAAGIVDPAGTGVANAGLVYFNGRLAMSDDLPIYQVQITPNGDLTKVGREDFDGOLES 292  
Db 238 GLFGLVDHSGKGTGVANAGLVYFNNRLAMSDDLPIYHVKTPTGDLTKTEGRDFDGOLES 297  
QY 293 TMIAHPKVPDPESGELFALSVDVSKPYLYKFRFPDGTGKSPDVEIQLDQTMHDFAIT 352  
Db 298 TMIAHPKLDPVSGELFALSVDVIQKPYLYKFRFPKNGEKSNDEIPEVEDPTMHDFAIT 357  
QY 353 NVVVVDDQVTFKLPDMIRGSGSPVYDKNVARGLDKYAEFDSNKNWIDAPDCFCFHL 412  
Db 358 NWVVPDQVQVFKMSEIRGSGSPVYDKNVARSFGLDKYAKGSDLKWVDPDCFCFHL 417  
QY 413 NNAWEPEETDEVVVIGSCMTTPDSTFNESDENLKSLSVSEIRLNLKTGSTRRPISNEDQ 472  
Db 418 NNAWEAEETDEIVVIGSCMTTPDSTFNECEDEGLKSVLSEIRLNLKTGKSTRKAIENPDE 477  
QY 473 QVNLEAGMVRNMLGRKTKFAYLALAEFPWPKVSGFAKVDLTITGEVKKHLYGDNRYGGEPL 532  
Db 478 QVNLEAGMVRNMLGRKTKFAYLALAEFPWPKVSGFAKVDLTITGEVKKHLYGDNRYGGEPL 537  
QY 533 FLPGS--GGEDEGYILCFVHDEKTKSELQIVNAVSLVEATVKLPSRVPYGFHGTFFIG 590

Db 538 FLPRDPNSKEEDDGYILAFVHDEKTKSELQIVNAVSLVEATVKLPSRVPYGFHGTFFIN 597  
QY 591 ADDLAKQ 597  
Db 598 ANDLANQ 604

RESULT 3  
E36812  
protein F399.10 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: E36812  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alons  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzial  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: E36812  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-657 <STO>  
A:Cross-references: GB:AB005173; NID:g8052533; PIDN:AAF71797.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F3F9.10  
A:Map position: 1

Query Match 68.9%; Score 2170.5; DB 2; Length 657;  
Best Local Similarity 67.7%; Pred. No. 1.6e-156;  
Matches 419; Conservative 80; Mismatches 87; Indels 33; Gaps 9;

QY 1 MASFATATAAAGGRLM-----GGNHPTPPSSSSQSSDLSSYCSLP-----MASRVT 45  
Db 51 MASITLLSTSTQFLDRFTSSSSSRPKQS-----LSPSTLRNKKLVPCVSSSVN 105  
QY 46 RKLNVSSALHT-----PPALHFFKQSSNSPAIVVVKPKAKESNTKQNLFORAAAAALDAE 101  
Db 106 KKSSVSSLSQSTFKPPSW---KKLCNDVTNLI-PKTTNQNP-K-LNPVQRTAAVMVDAVE 160  
QY 102 GLVLSHE-KLHPLPKTADPSVQIAGNAPVNEOPVRNLPVVKLPDSIKGVVVRNGANP 160  
Db 161 NAWISHEERRRPHPTADPAVQIAGNFPFVEKPVVHNLPTGIVPECIQGVVVRNGANP 220  
QY 161 LHEPVTGHHFFDGDGMVHAVKFEHGSASACRFTQNRFFVQERQLGRPVFPKPAIGELHGH 220  
Db 221 LHKPVSGHLLFDGDGMVHAVRFDNGSVSYACRFTETNRLVQERECGRFVFPKPAIGELHGH 280  
QY 221 TGIARLMLFYARAAAGIVDPAHGTGVANAGLVYFNGRLAMSDDLPIYQVQITPNGDLTK 280  
Db 281 LGIAKLMLFNTKGLFGLVDPTGGLGVANAGLVYFNGHLLAMSDDLPIYHVKTPTGDLTK 340  
QY 281 VGRFDPDGOLESTMIAPKVPDPESGELFALSVDVSKPYLYKFRFPDGTGKSPDVEIQLD 340  
Db 341 SGRYDFDGOLESTMIAPKVPDPESGELFALSVDVSKPYLYKFRFPDGTGKSPDVEIQLD 400  
QY 341 OPTMHDFAITNFVVPDQVQVFKLPDMIRGSGSPVYDKNVARGLDKYAEFGLDKYABSSNIK 400  
Db 401 OPTMHDFAITNFVVPDQVQVFKLPDMIRGSGSPVYDKNVARGLDKYAEFGLDKYABSSNIK 460  
QY 401 WIDAPDCFCFHLNNAWEPEETDEVVVIGSCMTTPDSTFNESDENLKSLSVSEIRLNLKTGE 460  
Db 461 WIEVPCFCFHLNNSWEPEETDEVVVIGSCMTTPDSTFNEHDETTLSVLSSEIRLNLKTGE 520  
QY 461 STRPPIISNEDQVNLNLEAGMVRNMLGRKTKFAYLALAEFPWPKVSGFAKVDLTITGEVKKH 520  
Db 521 STRPPIISNEDQVNLNLEAGMVRNMLGRKTKFAYLALAEFPWPKVSGFAKVDLTITGEVKKH 577

QY 521 LYGDNRYGGEPLFLPGEGEEDEGYILCFVHDEKTKWSELOIVNAVSLVLEATVKLPSRV 580  
 Db 578 IVGEGYKGEPLFLPGSGDEEGGYIMVPHBEKVSELOIVNAVNMKLEATVTLPSRV 637  
 QY 581 PYGFHGTFTGADDLAKQVV 599  
 Db 638 PYGFHGTFTSKEDLSKQAL 656

## RESULT 4

A86425  
 probable 9-cis-epoxycarotenoid dioxygenase [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: A86425  
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: A86425  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-589 <STO>  
 A:Cross-references: GB:AE005172; NID:gl1094779; PIDN:AAG29711.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1

Query Match 64.9%; Score 2044.5; DB 2; Length 589;  
 Best Local Similarity 68.1%; Pred. No. 4.9e-147;  
 Matches 390; Conservative 69; Mismatches 101; Indels 13; Gaps 4;

QY 27 SQSDLSYCSLPMASRVTKLVSSALHTPPALHPKQSSNSPAIVKPKAESNTKQ 86  
 Db 27 SPSSVSFTTKPR-----RKLSANSVSDTFNLLNFNYPSPNPII-----PEKDTSRW 76  
 QY 87 NLFORAAAALDAAGFLVSHKHLPLKTPADSPVQIAGNFAPVNEQPVVRNLPPVVGKLP 146  
 Db 77 NPLQRAASAALDAEATALLRRERSKPLPKTVDRHQISGNYAPVPEQSKSSLSVDGKIP 136  
 QY 147 DSITGVTVRNGANPLHEPVTHGHFFDGDGMVHAKVEHGSASVACRFTQTNRPVQRQLG 206  
 Db 137 DCIDGVYLRNGANFLEPVSCHLFDGDGMVHAKVITNGDASYSCTFTETRLVQSKQLG 196  
 QY 207 RPVPFKAIGELHGTGIARMLFYARAAAGIVDPAGTGVANAGLVYFNGRLIAMSDDL 266  
 Db 197 SPIFFKAIGELHGSHTARMLFYARCLFOLLNKHNGTGVANAGLVYFHDRLLANGSDDL 256  
 QY 267 PYQVQITPNGDLTKTVGRFDFDGOLESTMIAPKVDPESELPAISYDVVSKPKYLKFRPS 326  
 Db 257 PYQVRVTDNGDLETIGRDFDGOLESSAMIAHPKIDPVTKELFALSVDVVKPKYLKFRPS 316  
 QY 327 PDGFKSPDVEIQDQPTMHDEAITENFVVVPPQVVKLPEMIRGSPVYDKNKVARF 386  
 Db 317 PEGEKSPDVEIPLASPTMHMDPAITENFVPPQVVKLSMDFLGKSPVKVDGEKISR 376  
 QY 387 GILDKYAEDSSNIKWIDAPPCFCFHLNMAWEPETDEVVVGSCMTTPDPSIFNESDENLK 446  
 Db 377 GILPRNAKASAMVWVESPETFCFHLNMAWESPETDEVVVGSCMTPADSIFNECDQLN 436  
 QY 447 SVLSEIRLNKTKGSTRPPIISNEDQOVNLEAGVNRNMLGRKTKAYLALAPWPVKVG 506  
 Db 437 SVLSEIRLNKTKGSTRPPIIG-SVQMNLEAGVNRNMLGRKTRAYLALAPWPVKVG 495  
 QY 507 FAKVDLTGVEKHYLDNRYGGEPLPG--EGGEDEGYILCFVHDEKTKWSELOIVN 564  
 Db 496 FAKVDLTGVEYONHYFGKKGKYGGEPLPRGLESDDGDDGYIMSFVHDEBESWELHIVN 555

QY 565 AVSLEVEATVKLPSRVYPYGFHGTIGADDLAKQ 597  
 Db 556 AVTLELEATVKLPSRVYPYGFHGTIFVNSADMLNQ 588

## RESULT 5

T04531  
 nine-cis-epoxycarotenoid dioxygenase homolog F28J12.10 - Arabidopsis thaliana  
 N:Alternate names: hypothetical protein T9A21.200  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-Jun-1999  
 A:Accession: T04531; T04937  
 R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,  
 submitted to the Protein Sequence Database, February 1998  
 A:Reference number: Z15377  
 A:Accession: T04531  
 A:Molecule type: DNA  
 A:Residues: 1-583 <BEV>  
 A:Cross-references: EMBL:AL021710  
 A:Experimental source: cultivar Columbia; BAC clone F28J12  
 R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.P.  
 submitted to the Protein Sequence Database, February 1999  
 A:Reference number: Z15390  
 A:Accession: T04937  
 A:Molecule type: DNA  
 A:Residues: 1-377 <BSW>  
 A:Cross-references: EMBL:AL021713  
 A:Experimental source: cultivar Columbia; BAC clone T9A21  
 C:Genetics:  
 A:Map position: 4  
 A:Note: F28J12.10; T9A21.200

Query Match 63.2%; Score 1991; DB 2; Length 583;  
 Best Local Similarity 64.5%; Pred. No. 5.5e-143;  
 Matches 380; Conservative 82; Mismatches 105; Indels 22; Gaps 6;

QY 16 GGNETQPLSSQSSDLSY--CSSLPWASRVTKLVSSALHTPPALHPKQSSNSPAIV 73  
 Db 11 GGKITWP----QAQIDLDGFRPIKQPKVIKTQVIDYTE-LTKRQLFTRTTATTP-- 62  
 QY 74 VKPKAESNTKQMLFORAAAALDAAGFLVSHKHLPLKTPADSPVQIAGNFAPVNEQ 133  
 Db 63 -----QHNPLRLNIFOKAAAIAIDAAERALISHEQDSPLKTPADSPVQIAGNYPVPS 116  
 QY 134 PVBRNLPPVVKLPDSIKGVVVRNGANPLHEPVTHGHFFDGDGMVHAKVEHGSASVACRF 193  
 Db 117 SVRRNLTVGPTDPCIDGVYRNGANPNFETAGHLLFDGDGMVHAKVITNGSASVACRF 176  
 QY 194 TQTNRFVQERQLGRVPFPKAIAGELHGTGIARMLFYARAAAGIVDPAGTGVANAGLVY 253  
 Db 177 TKTERLVQEKRLGRFPVPFKAIGELHGSHTARMLFYARGLCLGGLNNGQGVANAGLVY 236  
 QY 254 ENGRLLAMSEDDLEYQVQITENGDLTKTVGRFDFDGOLESTMIAPKVDPESELPAISYD 313  
 Db 237 FNNRLAMSEDDLPYQKITQITGOLQTVGRYDFDGOQKASAMIAHPKIDPVTKELHALSYD 296  
 QY 314 VVSKPKLYKFRFSPDGTGKSPDVEIQDQPTMHMDPAITENFVVVPPQVVKLPEMIRGG 373  
 Db 297 VVKPKLYKFRFSPDGKVSPELEIPLPTMIHDFAITENFVTPDQVVPKLGEMISGK 356  
 QY 374 SPVYDKNKVARFGLDKYAEDSSNIKWIDAPPCFCFHLNMAWEPETDEVVVGSCWTP 433  
 Db 357 SPVYFDGKVSRLGIMPDKATEASQIIWVNSPETFCFHLNMAWESPETEIVVIGSCWSP 416  
 QY 434 PDSIFNESDENLKSVLSEIRLNKTKGSTRPPIISNEDQOVNLEAGVNRNMLGRKTKFA 493  
 Db 417 ADSIFNEDESLRSVLSEIRLNKTKGSTRPPIISNEDQOVNLEAGVNRNMLGRKTKFA 474  
 QY 494 YIALAEPWPVKSGFAKVDLTGVEKHYLDNRYGGEPLPGEGG-----EEDGYILCF 549  
 Db 475 FLATAYWPVKSGFAKVDLTGVEKHYLDNRYGGEPLPGEGG-----EEDGYILCF 534

Qy 550 VHDEKTKWSELOIWNVALEAVTKLPSRPVGFHGTFFGADDLAKOV 598  
Db 535 VHDSEKTSLOIWNVALEAVTKLPSRPVGFHGTFFGADDLAKOV 583  
RESULT 6  
T04351  
viviparous-14 protein - maize  
N:Alternate names: vpl4 protein  
C:Species: Zea mays (maize)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 08-Oct-1999  
C:Accession: T04351  
R: Tan, B.C.; McCarty, D.R.  
submitted to the EMBL Data Library, March 1997  
A:Reference number: 215304  
A:Accession: T04351  
A:Status: Preliminary; translated from GB/EMBL/DBDJB  
A:Molecule type: mRNA  
A:Residues: 1-604 <TAN>  
A:Cross-references: EMBL:U95953; NID:g2232016; PIDN:AAB62181.1; PID:g2232017  
A:Experimental source: strain W22  
C:Genetics:  
A:Gene: vpl4  
A:Map position: 1  
C:Function:  
A:Description: catalyzes oxidative cleavage of 9-cis-epoxy carotenoids to yield xanthoxi

Query Match 61.3%; Score 1930; DB 2; Length 604;  
Best Local Similarity 63.3%; Pred. No. 2.5e-138;  
Matches 375; Conservative 73; Mismatches 122; Indels 22; Gaps 8;  
Qy 22 PFLSSQSSDLSYCSSUPMARVTRKLVSSALHTP---PALHPPKSSNSPAIV--- 74  
Db 17 PARSARAS-----NSVRFSRAVSSVPPAECLQAPFKPVADLPAPSRPAIAVFGHA 71  
Qy 75 -KPAKESNTKQMLFQRAAALDA-AEGFLVS-HEKHLPLKPTADPSVQIAGNPAPVN 131  
Db 72 AAPKAECKQLMLFQRAAALDAEEGFAVNLPRHGLPSTADPAVQIAGNPAVG 131  
Qy 132 EQVRRNLVYVVKLPDSITKGVYVNGANPLHEPTVGHFPGDGMVHAKVFEHGS-A 190  
Db 132 ERPPVHELVPVGRIPFFIDGVYVNGANPCFDPVAGHGLFDGDMVHAKVFEHGS 191  
Qy 191 CRFTQTNRFVQERQGRVFPKATGELHGTGIALMLFYARAAGIVDPAGHGVANAG 250  
Db 192 CRFTETARLROERATGRVPFPAIGELHGHSGIALMLFYARAAGIVDPAGHGVANAG 251  
Qy 251 LVYFNGRLAWSDDLQYQVQITPFGDLTKTVGRFDQGLSTMIHAPKVPDPESGELFAL 310  
Db 252 LVYFNGRLAWSDDLQYVHVRVADGDLETVGRYDFDQGLGCAMIHAPKLPDPATGELHAL 311  
Qy 311 SYDVTSKPYLYKFRSPDGTSPDVEIQLDQPTMMHDFAITENFVVPDQOVVFKLEMI 370  
Db 312 SYDVTKRPLYKYFRPDGTSDDDVEIQLDQPTMMHDFAITENFVVPDQOVVFKLEMI 371  
Qy 371 RGSFVVDVKNKVARFGILDKYAESNSIKWIDAPDCFCFFLWNAAWEEPETDEVVVGSC 430  
Db 372 RGSFVVDVKNKVARFGILDKYAESNSIKWIDAPDCFCFFLWNAAWEEPETDEVVVGSC 431  
Qy 431 MTPPDSIFNESDENIKSVLSEIRLNTKGTSTRPIISNEDQVNLKAGMVRNMIKGT 490  
Db 432 MTPPDSIFNESDENIKSVLSEIRLNTKGTSTRPIISNEDQVNLKAGMVRNMIKGT 490  
Qy 491 KFAYLALAEPPWPKVSGFAKVDLTGTEVKKHLYGDNRYGGEPLFLPGEGE-----EBEGY 545  
Db 491 RYAYLALAEPPWPKVSGFAKVDLTGTEVKKHLYGDNRYGGEPLFLPGEGE-----EBEGY 550  
Qy 546 ILCFVHDEKTKWSELOIWNVALEAVTKLPSRPVGFHGTFFGADDLAKOV 597  
Db 551 VLTFFVHDERAGTSSELLVWNAADIRLEATVQLPSRPVGFHGTFFGADDLAKOV 602  
RESULT 7

T04438  
hypothetical protein T18B16.140 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 30-Apr-1999  
C:Accession: T04438  
R: Bevan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Ansorge, W.; Bancroft, I.; Mewes, H.  
submitted to the Protein Sequence Database, April 1998  
A:Reference number: 215359  
A:Accession: T04438  
A:Molecule type: DNA  
A:Residues: 1-595 <BEV>  
A:Cross-references: EMBL:AL021687  
A:Experimental source: cultivar Columbia; BAC clone T18B16  
C:Genetics:  
A:Map position: 4  
A:Note: T18B16.140  
Query Match 30.7%; Score 968; DB 2; Length 595;  
Best Local Similarity 36.7%; Pred. No. 2.5e-65;  
Matches 210; Conservative 106; Mismatches 220; Indels 36; Gaps 13;  
Qy 42 SRVTRKLVSSGALHTPPALHPPKQSSNSPAIVVVKPKAKESNTKQMLFQRAAALDAAE 101  
Db 42 SPITNPSDNDRNKKPTLH---NRTNHTLVSSPPKLRPEMTALALF---TTVEDVIN 94  
Qy 102 GFLVSHKHLPLKPTADPSVQIAGNPAPVNPQVRRNLVYV-GLKPSIKGVYVNGANP 160  
Db 95 TFIDP-----PSRPSVDPKHVLSDNFAVLDLPPTDCEIHLGTLFSLNGAYIRNGNP 149  
Qy 161 LHEPTVGHFPGDGMVHAKVFEHGSASYACRFQTNRVQERQGRVFPKALGELHGH 220  
Db 150 QFLPGPFLHFDGDMHAIKHGKATLCSRYKYKYVVEKQTGAFVMPVFSFGNGV 209  
Qy 221 T-GIARLMLFYARAAGIVDPAGHGVANAGLVFNGRLAWSDDLQYQVQITPFGDLK 279  
Db 210 TASVARGALTAARVLVTGQYNPVNGIGLANTSLAFFSNRLFALGESDLPYAVRLTESGDIE 269  
Qy 280 TVGRDFDQGLSTMIHAPKVPDPESGELFALSVDVSKPYLYKFRSPDGTSPDVEI-Q 338  
Db 270 TIGRDFDQGLKAMSTAHPTDPTIGTETAFRYGV-PPFLTYFRFDAGKQKQDVPFIS 328  
Qy 339 LDQPTMMHDFAITENFVVPDQOVVFK-----LPEMIRGSSPVVYVKNKVARFGILDKYAE 395  
Db 329 MTSPSFLHDFAITKHAIFAETQLGVRNMMLDLVLEGSSPVGTGDKTFRILGVIPKYAGD 388  
Qy 396 SSNIKWIDAPDCFCFFLWNAAWEEPETDEVVIGSCMTPTPDSIFNESDENIKSVLSEIRLN 455  
Db 389 ESEMKNFVPPGNNIIHAINAMDEDDGNSVLIAPNIMSIETLERMD-LVHALVEKVKID 447  
Qy 456 LKTGSTTRPIISNEDQVNLKAGMVRNMLGRKTKFAYLALAEPPWPKVSGFAKVDLTG 515  
Db 448 LVTGIVRRHPISAR-----NLDFAVINPAFLGRCSRYVYAAIGDPMKPSIGVYVVKLDVSKG 502  
Qy 516 E-----VKHLYGDNRYGGEPLFLPGEGE-----EBEGYILCFVHDEKTKWSELOIWNVA 566  
Db 503 DRDDCTVAIRMTYSGCYGGEFFVARDPGNPAEEDDGVVTVYVHDEVTGSKFLVMDAK 562  
Qy 567 S--LEVEATVQLPSRPVGFHGTFFGADDLAK 596  
Db 563 SPELEIVAAVRLPRVVPYGFHGLFVKESDLNK 594  
RESULT 8  
T49193  
neoxanthin cleavage enzyme ncl - Arabidopsis thaliana  
N:Alternate names: protein MAA21.150  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C:Accession: T49193  
R: Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Mewes, H.W.; Rudd, S.; Lemcke,  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z25018  
A:Accession: T49193

A;Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-538 <RIE>  
A;Cross-references: EMBL:AL163818; GSPDB:GN00061; ATSP:MAA21.150  
A;Experimental source: cultivar Columbia; BAC clone MAA21  
C;Genetics:  
A;Gene: ATSP:MAA21.150  
A;Map position: 3 96/2; 101/2; 138/3; 172/1; 193/1; 230/1; 291/3; 338/1; 411/2; 49  
A;Introns: 72/3; 96/2; 101/2; 138/3; 172/1; 193/1; 230/1; 291/3; 338/1; 411/2; 49

Query Match 29.8%; Score 938; DB 2; Length 538;  
Best Local Similarity 37.0%; Pred. No. 4.1e-63;  
Matches 207; Conservative 105; Mismatches 199; Indels 48; Gaps 13;

QY 64 KQSSNSPAIVVVKPK-AKESNTKQMLFORAAAAALDAAGFLVSHKHLPLKPTADPSVQ 122  
DB 4 KLSGSGSIISVHPRPSKGFSSKLLDLRLVVKLM-----HDSLEPLH-----Y 47

QY 123 IAGNFAPV-NEQPVRRNLFPVVKLPDSIKGVYVRNGANPLHEPVTGHFFDGDGMVHAVK 181  
DB 4 KLSGSGSIISVHPRPSKGFSSKLLDLRLVVKLM-----HDSLEPLH-----Y 47

QY 48 LSGNFAPIRDETTPVKDLFVHGFLPECLNGEFVRVGNPKFPAVAGYHWFDDGMIHGV 107

QY 182 FHGGSASYACRTQTNRFRVQERQOLGRPVFPKPAIGELHGTGIALMLFYARAAAGIVDPA 241  
DB 108 IKDGKATYVSRYVKTSLKQBEFFGAAPFK-IGDLKGFGLLVNVOQLRTKILNT 166

QY 242 HGTGVANAGLVYFNGRLAMSEDDLPYQVOITPNDGLKTVGRFDPDGOLESTMIAPKVD 301  
DB 167 YNGTANTALVYHGHKLLALQEADEKPVYIKVLEDGLQTLGIIDYDKRLTHSFTAPKVD 226

QY 302 PESGELFALSYDVVSKPYLYKFRFSPDGTKSDEVEIQLODPTMMHDFAITENFVVVDDQ 361  
DB 227 PVTGEMFTFGYS-HTPPYLTYRVISKDGIMHDPVPTITISEPIMHDFAITETTYAIFMDLP 285

QY 362 VVFKLPEMIRGGSPV-YDKNKVARFGILDKYAEDSSNIKWIDAPDCFCFHLNNAWEEPE 420  
DB 286 MHPREKMWKEMKMIYSPDTPKARFGLPRYAKDELMIKWFELPNCIFHNANAWEE-E 343

QY 421 TDEVVVIGSMTPPD-----SIFNESDENLKSVLSEIRLNKLTGSTRRPIISNEQQVN 475  
DB 344 EDEVLLITCRLENPDLDMVSGVKLEFNGLNEMRFNMKTGSASQKLSASA----- 398

QY 476 LEAGMVRNMLGRKTFAYLALAEPPKVSQFAKVD-----LTTGEVKKHLV-- 522  
DB 399 VDFPRINECTGKQKRVYVGTILDSIAKVTGIIFDLHAEAEKGMLEVGNGIKGIYDL 458

QY 523 GDNRYGGEPLFPGEGBEDEGYILCFVHDEKTKWSELOIVNAVSLVE--ATVKLPSRV 580  
DB 459 GEGRYGSEAIYVPRETAEDDGLYIFVHDENTGKSCVTVIDAKTWSAEPVAVVELPHRV 518

QY 581 PYGPHGTFTGADDLAKQVV 599  
DB 519 PYGFHALFVTEEQLEQTL 537

RESULT 9  
T51734  
neoxanthin cleavage enzyme [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 18-Aug-2000  
C;Accession: T51734  
R;Neill, S.J.; Burnett, E.C.; Desikan, R.; Hancock, J.T.  
J. Exp. Bot. 49, 1893-1894, 1998  
A;Title: Cloning of a wilt-responsive cDNA from an Arabidopsis thaliana suspension culture  
A;Reference number: Z24454  
A;Accession: T51734  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-538 <NEI>  
A;Cross-references: EMBL:AJ005813; PIDN:CAA06712.1  
A;Experimental source: cultivar Landsberg erecta  
C;Genetics:  
A;Gene: ncl

Query Match 29.7%; Score 937; DB 2; Length 538;  
Best Local Similarity 37.0%; Pred. No. 4.8e-63;  
Matches 207; Conservative 105; Mismatches 199; Indels 48; Gaps 13;

QY 64 KQSSNSPAIVVVKPK-AKESNTKQMLFORAAAAALDAAGFLVSHKHLPLKPTADPSVQ 122  
DB 4 KLSGSGSIISVHPRPSKGFSSKLLDLRLVVKLM-----HDSLEPLH-----Y 47

QY 123 IAGNFAPV-NEQPVRRNLFPVVKLPDSIKGVYVRNGANPLHEPVTGHFFDGDGMVHAVK 181  
DB 48 LSGNFAPIRDETTPVKDLFVHGFLPECLNGEFVRVGNPKFPAVAGYHWFDDGMIHGV 107

QY 182 FHGGSASYACRTQTNRFRVQERQOLGRPVFPKPAIGELHGTGIALMLFYARAAAGIVDPA 241  
DB 108 IKDGKATYVSRYVKTSLKQBEFFGAAPFK-IGDLKGFGLLVNVOQLRTKILNT 166

QY 242 HGTGVANAGLVYFNGRLAMSEDDLPYQVOITPNDGLKTVGRFDPDGOLESTMIAPKVD 301  
DB 167 YNGTANTALVYHGHKLLALQEADEKPVYIKVLEDGLQTLGIIDYDKRLTHSFTAPKVD 226

QY 302 PESGELFALSYDVVSKPYLYKFRFSPDGTKSDEVEIQLODPTMMHDFAITENFVVVDDQ 361  
DB 227 PVTGEMFTFGYS-HTPPYLTYRVISKDGIMHDPVPTITISEPIMHDFAITETTYAIFMDLP 285

QY 362 VVFKLPEMIRGGSPV-YDKNKVARFGILDKYAEDSSNIKWIDAPDCFCFHLNNAWEEPE 420  
DB 286 MHPREKMWKEMKMIYSPDTPKARFGLPRYAKDELMIKWFELPNCIFHNANAWEE-E 343

QY 421 TDEVVVIGSMTPPD-----SIFNESDENLKSVLSEIRLNKLTGSTRRPIISNEQQVN 475  
DB 344 EDEVLLITCRLENPDLDMVSGVKLEFNGLNEMRFNMKTGSASQKLSASA----- 398

QY 476 LEAGMVRNMLGRKTFAYLALAEPPKVSQFAKVD-----LTTGEVKKHLV-- 522  
DB 399 VDFPRINECTGKQKRVYVGTILDSIAKVTGIIFDLHAEAEKGMLEVGNGIKGIYDL 458

QY 523 GDNRYGGEPLFPGEGBEDEGYILCFVHDEKTKWSELOIVNAVSLVE--ATVKLPSRV 580  
DB 459 GEGRYGSEAIYVPRETAEDDGLYIFVHDENTGKSCVTVIDAKTWSAEPVAVVELPHRV 518

QY 581 PYGPHGTFTGADDLAKQVV 599  
DB 519 PYGFHALFVTEEQLEQTL 537

RESULT 10  
AG1944  
hypothetical protein all1106 [imported] - Nostoc sp. (strain PCC 7120)  
C;Species: Nostoc sp. PCC 7120  
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C;Accession: AG1944  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, D.  
DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AG1944  
A;Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-475 <KUR>  
A;Cross-references: GB:BA000019; PIDN:BAB73063.1; PID:gl7130452; GSPDB:GN00179  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: all1106

Query Match 23.7%; Score 745.5; DB 2; Length 475;  
Best Local Similarity 34.5%; Pred. No. 1.4e-48;  
Matches 169; Conservative 91; Mismatches 171; Indels 59; Gaps 15;

QY 123 IAGNFAPVNEQPVRRNLFPVVKLPDSIKGVYVRNGANPLHEPVTGHFFDGDGMVHAVK 182  
DB 123 IAGNFAPVNEQPVRRNLFPVVKLPDSIKGVYVRNGANPLHEPVTGHFFDGDGMVHAVK 182

Db 22 LEQNFAPVBBEITDTTLKVIQELPPELSGMFVRNPNPQWTPIGOYHWFQDGMHLGVRI 81  
QY 183 EHSASVACRPTQNRVQEROLGRVPFPAKALGELHGHGTGIALMLFYARAAAGIVD--- 239  
Db 82 SNGKATYNRVVRQWQIEHQAQW-----TGLMEPPQ 117  
QY 240 ---PAHGTGVANAGLVFNGRLLAMSEDDLPYQVQITPNDGLKTVGRFDFDQLESTMLIA 296  
Db 118 TELSPRNTG--NTALIWHAGOLLALWEGGAPYATQVP---DLASIGETTYNNQLSSAFTA 172  
QY 297 HPKVDPSGELFALSVDVSKPYLKFRFSPDGTKSPDVEIQDQPTMMHDFALTENFVV 356  
Db 173 HPKVDPTGEMWFGYS--FAPYLYHSVVSATGELVTVFIDLPWGMHDFALTANYTI 231  
QY 357 VPDQOVFKLPEMIRGGSPVYVYDKNKVARFOILKYAEDSSNKKWIDAPCFCHLWNAW 416  
Db 232 FMDLPLATFVSVMORGEBMLMFESDRSPRFGILPRHG--DNSQIRWFEPASCYVFHTLNAY 230  
QY 417 EEPETDEVVIG-----SCTMPDPSIFNESDENLK--SVLSEIRLNKLTGSTRRPIISN 469  
Db 291 E--DKDEVVLACPMRSTTVLASDS---QTDPEADIPRLHRWFHLTKGVHEEML--- 342  
QY 470 EDQOVNLEAGMVRNMLGRKTKFAYLA--LAP--PWPKVSGFAKVDLTGTGEVKKHLYGDNRY 527  
Db 343 --DVASEFPRIENFLQPTQYGYTSRLAKGSIPLEGLIKYDLSNAKSQNYEYGGRY 400  
QY 528 GGEPLFIPGEGE--EDEGYILCFVHDEKTSSELOIVNAVSLVZ--ATVKLPSRVYGF 584  
Db 401 GSEAVFVRPRGATVEDDGLNLYVYDTEESSELVIVINAQDINSEPIARVLLPORVYGF 460  
QY 585 HGTFIGADDL 594  
Db 461 HGIWVTEBQL 470

RESULT 11  
T17019  
Hypothetical protein - apple tree  
C:Species: Malus domestica (apple tree)  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jul-2000  
C:Accession: T17019  
R:Watillon, B.; Kettmann, R.; Aredouani, A.; Hecquet, J.; Boxus, P.; Burny, A.  
Plant Mol. Biol. 36, 909-915, 1998  
A:Title: Apple messenger RNAs related to bacterial lignostilbene dioxygenase and plant  
A:Reference number: Z18655; MUID:98179104; PMID:9520281  
A:Accession: T17019  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-446 <WAT>  
A:Cross-references: EMBL:293765; NID:G2924324; PIDN:CA807784.1; PID:G2924325  
A:Experimental source: cv. McIntosh, strain Wjczk

Query Match 23.7%; Score 745; DB 2; Length 446;  
Best Local Similarity 39.8%; Pred. No. 1.3e-48;  
Matches 150; Conservative 76; Mismatches 129; Indels 22; Gaps 9;

QY 231 ASAAAGIVDPAGTGVANAGLVFNGRLLAMSEDDLPYQVQITPNDGLKTVGRFDFDGL 290  
Db 76 ASVLTGQYNPANGIGLANTSLAFCGRLYALGESDLPYSLRSLTNGDIETLGRHDFDGL 135  
QY 291 ESTMTAHKVPDSESELFALSVDVSKPYLKFRFSPDGTKSPDVEI--QLDQPTMMHDF 349  
Db 136 SNMNTAHPKIDPTGAEAFR'YGF1-RPLTYFRFDSNGVKQPDVPISWVTPPTFLHDF 194  
QY 350 ITENFVVVDDQOVFKLPEMI--RGGSPVYDKNKVARFGILDKYAEDSSNKKWIDAPCF 408  
Db 195 ITKGAIFAADIQIGLNLDIMTKRATPFLDPSKVPRIQVILYAKDESEMRWFEPGFN 254  
QY 409 CFHLNNAWEPETDEVVV'GSCWTPDPSIFNESDENLKSLSSEIRLNKLTGSTRRPIIS 468  
Db 255 GVHATNAND--EDDAIVWVAPNVLSEHVLERVD--LVHCLVEKVIDLKTGIVTEQLST 311  
QY 469 NEDQOVNLEAGMVRNMLGRKTKFAYLAEPWPKVSGFAKVDLTGTG-----VKKHLYG 523

Db 312 R-----NLDFAVINPAYLGRNKKVYAAEGDPMKISGVVVKLDVSNVEHKECIVASRMFG 366  
QY 524 DNRYGGEPLFLPGE---GGEDEGYILCFVHDEKTSSELOIVNAVSLVZ--LEVEATVKLP 577  
Db 367 PCYIGGEFFVAREPENPEADEDNGFLVSYVHDEKAGESRFLVMDAKSPQLDIIAAVRMP 426  
QY 578 SRVYGFHGTFIGADDL 594  
Db 427 RRVYGFHGLFVRESDL 443

RESULT 12  
E87345  
conserved hypothetical protein CC0776 [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: E87345  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kol  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete genome sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: E87345  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-483 <STO>  
A:Cross-references: GB:AE005673; NID:gl3422015; PIDN:AAK22761.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC0776

Query Match 16.8%; Score 528.5; DB 2; Length 483;  
Best Local Similarity 30.9%; Pred. No. 4.1e-32;  
Matches 160; Conservative 89; Mismatches 203; Indels 65; Gaps 22;

QY 108 EKLHLPKTDPS--VOIAGNFAPVNEQPVNRNLVW--GKLPDSIKGVVVRNGANPLHEP 164  
Db 4 ERLPPVRLSLGTNHPYMTGPTPQHEEVNANWDLVLEGAIPADLDGYLRNTENPVHDP 63  
QY 165 VTGHFFDGDGMVHAKVFEHGSAYACRFTOTNRFVQBRQLGRPVFPKAIQELHGHGTGIA 224  
Db 64 IGRYHFFDGDGMHIEFKGGAATYRNFVTRCFEAEQVNEGLW---GGLMDGPGVS 119  
QY 225 RLMLFYARAAAGIVDPAGT--GVANAGLVFNGRLLAMSEDDLPYQVQITPNDGL 278  
Db 120 K-----RFGFG---AFGALKDSASTDIVVHNGEAIATP-----YQGEAYRDLPL-TL 163  
QY 279 KTVGRFDFDQLESTMTIAHPKVDPSGELFALSVDVWSK--PYLKFRFSPDGTKSPDVE 336  
Db 164 ENLGVASW-APLEG-VSAHPKYDEATGELMFPNY---SKANPYMHYGVVDPGKRKYQG 218  
QY 337 IQLDQPTMMHDFALTENFVVVDDQOVFKLPEMIRGGSPVYDKNKVARFGILDKYAEDS 396  
Db 219 VPLPGPRLPHDVAFSKYAILNCLFVFWDQELMARDIHAVRLHKGIPSRFALVPR---EG 275  
QY 397 SNIKWIDAPDCFCFLHNAWEPETDEVV'IGSCWTPP-----DSIFNESDEN 444  
Db 276 GEPRWEAEPTVYLWLNAYE--DGDEVLDGYFOEKEIPRPLEGAPDGHGLHMLAYLDEH 333  
QY 445 LKSVLSEI---RLNKLTGSTRRPIISNEDQOVNLEAGMVRNMLGRKTKFAYLAEP- 500  
Db 334 --SFLPKLHRWRPNLTKGTETTEKHL-----DDRVL-LEFGMFKQYAGKYRYAISTTAKPG 386  
QY 501 WPKVSGFAKVDLTGTGEVKKHLYGDNRYGGER'FLPQEGG--EDEGYILCFVHDEKTSSE 559  
Db 387 WFLNFGVFKHLEJTGESNIALPEGYASEAPFAPKVGAVDEDDGVLVSFIIDENKAGE 446  
QY 560 LOIVNAVSLVEEATVKLPSRVYGFHGTFIGADDLAK 596  
Db 447 CLIVDAKREVEVCRIALPHKLSGCTHVWAGREMLTK 483





```
Query Match      12.5%; Score 394; DB 2; Length 490;
Best Local Similarity 25.7%; Pred. No. 6.8e-22;
Matches 138; Conservative 93; Mismatches 190; Indels 116; Gaps 27;

Qy 114 PKTADPSVQIAGNAP-----VNEQVRRNL---PVGKLPDSIKGVYVRNGA-----158
Db 5 PPTSSPSQR---SYSPQWLGRYQSQPQEWYDVEDVEGSIPPDLOQTLYRNGPGLLEIG 61

Qy 159 -NPLHEBVTGHHRPDGMVHAKFE-HGSASVACRFTQNRVQERQLGRPVPEPKAIGE 216
Db 62 DRPLKHP-----FDGGMWTAFAFGDGRVHFQSKFVRTQGVBEQKAGKMIYRGVFGS 115

Qy 217 LHGTGIARLMLFYARAAAGIVDPAGH-----TGVANAGLVYFNGRLAMSEDDL 266
Db 116 -----QAGGWLKTIFDLRLKNIANTNITWGDRLIALWEGGQ 153

Qy 267 PYQVQITENGDKTVGRFPDQGLE--STMIAHPKVDPE-----GELFALSYDVVS--KP 318
Db 154 PHRL-----PSNLATIGDDLLGGLAECQPLSAHPRIDPASTFDGGQPCYVTFSIKSLSS 210

Qy 319 YLKYFRFSPDGTSFVVEIQLDQFT-----MMHDFAITENFVYVDPQQVFEK-LPEM- 369
Db 211 TLTLELDPOGK-----LLRQKTETPGFAFIHDFAITPHVAIFLQNNVTNLGLPYLF 263

Qy 370 -IRG-GSPVYDKMKVAFGLDKYAEDESSNIKWIDAPDCFCFHLNNAWEEPTDEWVI 427
Db 264 GLRGAGECVQHPDKPAQIILVPR---DGBEIKRIPVOAGFVFHANAFAE--ENGKIILD 318

Qy 428 GSCMTFPPDSIFNESDENLKSVLSEIRLNLTGSTR---RPIISNEDQQVNL----EAGM 480
Db 319 SICYNSLPQV--DIDGDFRSTNFD--NLDPGQLWRFTIDPAAATVEKQLMWSRCCEFPV 373

Qy 481 VNRNMLGRKTKFAYLALAE-----PWPKVSGFAKVDITTEGVKKHLYGDNRYGGBPLFL 534
Db 374 VHPQOVGRPYRYVYNGAAHSTGNAP---LQAILKVDLESCTETLRSFAPHGAGEPIFV 430

Qy 535 PEGGG-EEDEGVILCFVHDEXTKSELQIVNA--VSLEVEATVKLPSRPVYGPHGTF 588
Db 431 PRPGVAEDDGMLLCLIIYKADHRSELVILDAQDITAPATATLKLKHIPYPLHGSW 487
```

Search completed: November 14, 2003, 21:19:12  
Job time : 23 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 16, 2003, 22:18:27 ; Search time 516 Seconds

(without alignments)  
3793.904 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150

Sequence: 1 MASFTATAVSGRWLGNGHT.....VPYGHGTFIGADDLAKQVV 599

Scoring table:

BLOSUM62					
Xgapop 10.0	Xgapext 0.5				
Ygapop 10.0	Ygapext 0.5				
Fgapop 6.0	Fgapext 7.0				
Dgapop 6.0	Dgapext 7.0				

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ .p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09758269/runat\_14112003.192311.25851/app\_query.fasta.1.775  
-DB=Published Applications NA -QWMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORX=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 USER=US09758269 @CGN 1 1 347 @runat 14112003 192311 25851  
-NCFU=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2:	/cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3:	/cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5:	/cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6:	/cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7:	/cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8:	/cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9:	/cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10:	/cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11:	/cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12:	/cgn2_6/ptodata/1/pubpna/US09D_PUBCOMB.seq:
13:	/cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14:	/cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15:	/cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16:	/cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
17:	/cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	3150	100.0	1800	10	US-09-758-269-5

2	2380.5	72.4	1818	10	US-09-758-269-15
3	2168.5	68.8	1839	10	US-09-758-269-11
4	1991	63.2	1752	10	US-09-758-269-1
5	1930	61.3	1815	10	US-09-758-269-13
6	1663.5	52.8	1734	10	US-09-758-269-9
7	969	30.8	1788	10	US-09-758-269-3
8	969	30.8	1788	10	US-09-938-842A-1444
9	937	29.7	1617	10	US-09-758-269-7
10	937	29.7	1617	10	US-09-758-269-17
11	522	16.6	393	10	US-09-878-574-2543
12	419	13.3	32679	8	US-08-976-0630-21
13	411.5	13.1	1518	8	US-08-976-0630-21
14	369	11.7	320	9	US-09-878-574-2872
15	346	11.0	230	9	US-09-770-696-172
16	279	8.9	534	12	US-10-149-758-65
17	260	8.3	3111	14	US-10-053-192-2
18	238.5	7.6	2134	12	US-10-168-517-18
19	234.5	7.4	367	9	US-09-770-751-402
20	232.5	7.4	1855	12	US-10-168-517-16
21	207.5	6.6	1934	12	US-10-168-517-20
22	150	4.8	2037	12	US-10-168-517-1
23	133	4.2	82993	15	US-10-080-170-645
24	127.5	4.0	13029	9	US-09-815-242-4052
25	122	3.9	5127	14	US-10-156-761-3630
26	122	3.9	9025608	14	US-10-156-761-1
27	111.5	3.5	1875	14	US-10-156-761-3143
28	111	3.5	1956	14	US-10-156-761-4583
29	110	3.5	9025608	14	US-10-156-761-1
30	108.5	3.4	4146	12	US-10-032-585-6784
31	107	3.4	2825	12	US-10-289-757-15
32	107	3.4	3075	9	US-09-833-790-424
33	107	3.4	7515	12	US-10-101-510-202
34	107	3.4	7515	12	US-10-210-120-4
35	107	3.4	81940	12	US-09-759-508B-1
36	107	3.4	81940	12	US-09-873-319-724
37	107	3.4	81940	12	US-09-360-706-1092
38	106.5	3.4	1083	9	US-09-815-242-7396
39	106.5	3.4	10174	14	US-10-171-311-82
40	106	3.4	1239	10	US-09-922-261-76
41	106	3.4	1825	10	US-09-922-261-75
42	106	3.4	2514	14	US-10-156-761-2792
43	106	3.4	4308	12	US-10-102-549-1
44	106	3.4	4331	12	US-10-354-358-67
45	105.5	3.3	3048	14	US-10-156-761-3146

ALIGNMENTS

RESULT 1

US-09-758-269-5

Sequence 5, Application US/09758269

Patent No. US20020104120A1

GENERAL INFORMATION:

APPLICANT: IUCHI, SATOSHI

APPLICANT: KOBAYASHI, MASATOMO

APPLICANT: SHINOZAKI, KAZUO

TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN

TITLE OF INVENTION: CLEAVAGE ENZYME GENE

FILE REFERENCE: 3914-3

CURRENT APPLICATION NUMBER: US/09/758,269

CURRENT FILING DATE: 2001-01-12

PRIOR APPLICATION NUMBER: JP 2001-003476

PRIOR FILING DATE: 2001-01-11

PRIOR APPLICATION NUMBER: JP 2000-010056

PRIOR FILING DATE: 2000-01-13

NUMBER OF SEQ ID NOS: 33

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 5

LENGTH: 1800

TYPE: DNA

ORGANISM: Arabidopsis thaliana

FEATURE:

NAME/KEY: CDS

; LOCATION: (1)..(1797)  
US-09-758-269-5

## Alignment Scores:

Pred. No.: 0 Length: 1800  
Score: 3150.00 Matches: 599  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 10 Gaps: 0

US-09-758-269-6 (1-599) x US-09-758-269-5 (1-1800)

```
QY 1 MetAlaSerPheThrAlaThrAlaAlaValSerGlyArgTrpLeuGlyGlyAsnHisThr 20
Db 1 ATGGCTTCTTACGGCAACGGCTGCGGTTCTGGAGATGGCTGGTGGCAATCACT 60

QY 21 GlnProProLeuSerSerGlnSerSerAspLeuSerTyrCysSerSerLeuProMet 40
Db 61 CAGCGCGCATATCGCTTCTTCAAGCTCCGACTTGAGTTATTGTAGCTCCTTACCTATG 120

QY 41 AlaSerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProAlaLeu 60
Db 121 GCCAGTCGTGCACACGTAAAGTCAATGTTTCATCTCGCTTCACACTCCCTCCAGCTCTT 180

QY 61 HisPheProLysGlnSerSerAsnSerProAlaIleValValLysProLysAlaLysGlu 80
Db 181 CATTTCCCTAAGCAATCATCAAACTCTCCGCCCAATGTTGTTAAGCCCAAGCCAAAGAA 240

QY 81 SerAsnThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaAlaLeuAspAlaAla 100
Db 241 TCCAACTAAACAGATGAATTTGTTCCAGAGAGCGCGCGGCGAGCGTTGGACGCGCG 300

QY 101 GluGlyPheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSer 120
Db 301 GAGGGTTTCCCTGTGTCAGCCACGAGAGCTACCCCGCTTCTCTAAACGGCTGATCTCTAGT 360

QY 121 ValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuPro 140
Db 361 GTTCAGATCGCCGGAATTTTGCTCCGGTGAATGAACAGCCGCTCCGCGCTAATCTCCG 420

QY 141 ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnPro 160
Db 421 GTGGTCGGAACACTTCCCGATCCATCAAGAGGTGTATGCGCAACGAGGTAAACCCA 480

QY 161 LeuHisGluProValThrGlyHisPheAspGlyAspGlyMetValHisAlaVal 180
Db 481 CTTTCAGACCGCGTGACAGGTCAACACTTCTTCGACGAGACGCTATGTTTCACGCGTC 540

QY 181 LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal 200
Db 541 AAATTTCGAACACGGTTCAGTACGCTTACCGCTTACCTCAAGCCATCGGTGASCTTCA 600

QY 201 GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHis 220
Db 601 CAGGAACGTCANTGGGTGACCGGTTTCCCAAGCCATCGGTGASCTTCAAGCCGAC 660

QY 221 ThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAspPro 240
Db 661 ACCGGTATTGCCGACTCATGCTATTCTACGCGCAGAGCTGCGCGGTATAGTCGACCG 720

QY 241 AlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAla 260
Db 721 GCACCGGAACCGGTAGCTAACCGCGGTTTGGTCTATTTCATTCAGCGCGGTTATTGGCT 780

QY 261 MetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThr 280
Db 781 ATGTCGGAGATGATTACCTTACCAAGTTCAGATCACTCCCAATGGAGATTAAAAACC 840

QY 281 ValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysVal 300
Db 841 GTTGGTGGGTTCGATTITGATGACAAATTAGAAATCCCAATGATGCCCAACCCGAAAGTC 900
```

```
QY 301 AspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeu 320
Db 901 GACCCGAATCCGGTGAACCTTTCGCTTAAAGTAGACGCTGCTTTTCAAGACCTTACCTA 960

QY 321 LysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAsp 340
Db 961 AAATACCTCCGATTCTCACCGGACGGAATAAATCAACCGACGTCGAGATTCAGCTTAT 1020

QY 341 GlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValValProAspGln 360
Db 1021 CAGCAACAGATGATGCACGATTTCGCGATTACAGAGAACTTCGTCGTCGACCTGACAG 1080

QY 361 GlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyrAspLys 380
Db 1081 CAATGCTGTTTCAAGCTCCGAGATGATCATCGCGTGGTCTCCGGTGGTTTACGACAAG 1140

QY 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnLys 400
Db 1141 AACAGGTCGCAAGATTTCGGGATTTCAGCAAAATACCGCAAGATTTCATCGAATTAAG 1200

QY 401 TrpIleAspAlaProAspCysPheHisLeuTrpAsnAlaTrpGluGluProGlu 420
Db 1201 TGGATTGATGCTCAGATTGCTTCTGCTTCATCTCTGGAACGCTTGGAGAGACCGAA 1260

QY 421 ThrAspGluValValIleGlySerCysMetThrProProAspSerIlePheAsnGlu 440
Db 1261 ACAGATGAAGTCGTCGATAGGCTCTGTATGATCCTCCACGACTCAATTTTCAACGAG 1320

QY 441 SerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGlu 460
Db 1321 TCTGACGAGATCTCAAGAGTGTCTGCTCGAATCCGCTGAATCTCAAAACCGGTGAA 1380

QY 461 SerThrArgArgProIleIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMet 480
Db 1381 TCAACTCGCGCTCCGATCATCTCCAACGAAGATCAACAAAGTCAACCTCGAAGCAGGATG 1440

QY 481 ValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluPro 500
Db 1441 GTCAACAGAAACATGCTCGCCGCTAAACCAATTCGCTTACTTGGCTTTAGCCGAGCG 1500

QY 501 TrpProLysValSerGlyPheAlaLysValAspLeuThrGlyGluValLysLysHis 520
Db 1501 TGGCTTAAAGTCTCAGGATTCGTAAGTTCATCTCACTTCTGGAGAAAGTTTAAAGAACAT 1560

QY 521 LeuTyrGlyAspAsnArgTyrGlyGluProLeuPheLeuProGlyGluGlyGlyGlu 540
Db 1561 CTTTACGGCGATAACCGTTTACGAGAGAGAGCTCTGTTCTCCCGGAGAGGAGAGAG 1620

QY 541 GluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLysSerGluLeu 560
Db 1621 GAAGACGAAGGATACATCCTCTGTTTCGTTTCAGCAGAGAGACATGGAATTCGAGTTA 1680

QY 561 GlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArgVal 580
Db 1681 CAGATAGTTAACGCCCTTAGCTTAGAGTTGAAGCAACCGTTAACTTCCGTCAGGGTT 1740

QY 581 ProTyrGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGlnValVal 599
Db 1741 CCGTACGATTTTCAACGCTACATTCATCGAGCGCATGATTTGCGGAAGCAGGTCGTG 1797
```

## RESULT 2

US-09-758-269-15  
; Sequence 15, Application US/09758269  
; Patent No. US20020104120A1  
; GENERAL INFORMATION:  
; APPLICANT: IUCHI SATOSHI  
; APPLICANT: KOBAYASHI, YASUTOMO  
; APPLICANT: SHINOZAKI, KAZUO  
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN  
; FILE REFERENCE: 3914-3  
; CURRENT APPLICATION NUMBER: US/09/758,269  
; CURRENT FILING DATE: 2001-01-12

Mon Nov 17 10:01:01 2003

```

, PRIOR APPLICATION NUMBER: JP 2001-003476
, PRIOR FILING DATE: 2001-01-11
, PRIOR APPLICATION NUMBER: JP 2000-010056
, PRIOR FILING DATE: 2000-01-13
, NUMBER OF SEQ ID NOS: 33
, SOFTWARE: Patent Ver. 2.1

```

```

/ SEQ ID NO: 1
/ LENGTH: 1818
/ TYPE: DNA
/ ORGANISM: Lycopersicon esculentum
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(1815)
US-09-758-269-15

```

Alignment Scores:	4,34e-252	Length:	1318
Pred. No.:	2280.50	Matches:	435
Score:	82.04%	Conservative:	63
Percent Similarity:	71.66%	Mismatches:	90
Best Local Similarity:	72.40%	Indels:	19
Query Match:	10	Gaps:	6
DB:			

US-09-758-269-6 (1-599) x US-09-758-269-15 (1-1818)

QY	5	ThrAlaThrAlaAlaValSerGlyArgTrpLeuGlyGlyAsnHisThrGlnProProLeu	24
DB	6	...:::...	
DB	7	ACTACTTTCACATGCCACAATAACATGAGT-----AAGACTAAGTTG	51
QY	25	SerSerSerGlnSerSerAspLeuSerTyrCyaser-----SerLeuProMeta	41
DB	52	TCATGCGCATCATCAAGGAGTTGGTTTGGCATCAAACTCTATTTCCTCTACTCAAAAT	111
QY	42	SerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProProAlaLeuHis	61
DB	112	...:::...	
DB	112	CAACATAATAGCCAAAGTCTCAACATTAATTCCTCTCTTCAAGCTCCACTATACTTCA	171
QY	62	PheProLysGlnSerSerAsn-----SerProAla-----IleValVallysProLys	77
DB	172	TTTCTCTAACCAATCTTCAAAATATCAACACCAAGAATAATACATTTTCACACCCANA	231
QY	78	AlaLysGluSerAsn-----ThrLysGlnMetAsnLeuPheGlnArgAla	92
DB	232	CAAGAAAAACAACAACCTCTCTTCTTCTTCAACTTCCAAAGTGGAAATTTAGTGCGAAGA	291
QY	93	AlaAlaAlaAlaLeuAspAlaAlaGluGlyPheLeuValSerHisGluLysLeuHisPro	112
DB	292	GCGAGCAATGGCTTTAGATGCTGTAGAAAGTCTTTAACTAAACATGAACCTTGAACACCC	351
QY	113	LeuProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAlaProValAsnGlu	132
DB	352	TTGCGGAAACAACGCCAGCCAGTCCAGATTTCTGGGAATTTTGTCCGGTACCGGAA	411
QY	133	GlnProValArgArgAsnLeuProValValGlyLysLeuProAspSerIleLysGlyVal	152
DB	412	AATCCAGTCTGTCAATCTCTTCGGGTACCGGAAAAATCCAAATGTGTTCAGGCGGT	471
QY	153	TyrValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHisHisPhePheAsp	172
DB	472	TAGTTCGAAACGGAGCTAACCCCTCTTTTGAACCAACCGCGGACACCATTTCTTCGAC	531
QY	173	GlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerTyrAlaCysArg	192
DB	532	GCGCAGCGGTATGGTTACACCGCTTCAATTCAAAAATGGGTGGGTAGTTACGCTTCCGCT	591
QY	193	PheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPheProLys	212
DB	592	TTCACTGAACACAGAGGCTTGTTCAGAAAAAAGCTTTGGTTCGCCCTGTTCCTTAA	651
QY	213	AlaIleGlyLeuLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyrAlaArg	232
DB	652	GCCATTGGTGAATTACATGCTCACTCGAATTCAGAGCTTATGCTGTTTACGCTCGT	711

Db 1792 GCCAATGATTGGCAAAATCAG 1812  
|||||::|||::|||::|||

RESULT 3  
US-09-758-269-11  
; Sequence 11, Application US/09758269  
; Patent No. US20020104120A1  
; GENERAL INFORMATION:  
; APPLICANT: IUCHI, SATOSHI  
; APPLICANT: KOBAYASHI, NASATOMO  
; APPLICANT: SHINOZAKI, KAZUO  
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN  
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE  
; FILE REFERENCE: 3914-3  
; CURRENT APPLICATION NUMBER: US/09/758,269  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: JP 2001-003476  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: JP 2000-010056  
; PRIOR FILING DATE: 2000-01-13  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 11  
; LENGTH: 1839  
; TYPE: DNA  
; ORGANISM: Vigna unguiculata  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1836)  
US-09-758-269-11

Alignment Scores:  
Pred. NO.: 3,53e-239 Length: 1839  
Score: 2168.50 Matches: 423  
Percent Similarity: 78.44% Conservative: 61  
Best Local Similarity: 68.56% Mismatches: 98  
Query Match: 68.94% Indels: 35  
DB: 8 Gaps: 8

US-09-758-269-6 (1-593) x US-09-758-269-11 (1-1839)

	Qy	8 AlalaValSerGlyArgTrpLeuGlyGlyAsnHisThrGlnProPheLeuSerSerSer	Ser	27
Dy	7 TCATCAGCTTCAAACACTTGTTTAACGCCACACTCCCATCTCCCCCTTCAAGACACTA	66		
Qy	28 GlnSerSerAspLeuSerTyrCysSerSerLeuProMetAlaSerArgValThrArgLys	47		
Dy	67 CTTCACCATCT--TCCTCCACAACATCTCTCTTTA-----AGAAAAACATCTCT	117		
Qy	48 LeuAsn---ValSerSerAlaLeuHisThrProProAlaLeuHisPheProLysGln---	65		
Dy	118 TCCAACACCATCACATGTCCCTTCAAACA-----CTCCACTTCCCCAACACAGTAC	169		
Qy	66 -----SerSerAsnSerProAlaIleValLysProLysAlaLysGluSer	81		
Dy	169 CAACCAACATCCATCCATCCACAGCCACCACCAACACCCCAATCAAACT	228		
Qy	82 AsnThr-----	83		
Dy	229 ACCACCATCACACCACCAACCGGCCAGGAACCAACCTCTCTGCACACCAACCA	288		
Qy	84 -----LysGlnMetAsnLeuPheGlnAAGAlaAlaAlaLeuAspAlaAla	100		
Dy	289 CCATTACCTCAAAAATGGAACCTTCTCCAGAAGCGCTGCCAGCGCTTGACCTGGTC	348		
Qy	101 GluGlyPheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaaspProSer	120		
Dy	349 GAACCGGCGCTCGTCTCGCACGAGCGCAACACCGCTCCCCAAAACGGGACCCGAGG	408		
Qy	121 ValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuPro	140		
Dy	409 GTCCAAATCCCGGGAACCTCGCGCGGTGGCGGAGCATGCCCGCATCAAGGACTCCCG	468		

Db 1546 TGGCCCAAGTCTCGGGCTTCGGAAGTGTGTTGCTGAGTGGGGAAGTGAAGAGTAC 1605  
Qy 521 LeuTyrGlyAspAsnArgTyrGlyGluProLeuPheLeuProGlyGluGlyGlyGlu 540  
Db 1606 ATGTATGAGAGAGAGAGATCGTGGGAGCCTCTGTTCTTCC--AAAGGCCAAAAA 1662  
Qy 541 GluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTyrPheSerGluLeu 560  
Db 1663 GAAGACGATGGGTATATTCGTCATCGTGCATCGTGCACGACGAGAAAGATCGAGCTG 1722  
Qy 561 GlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArgVal 580  
Db 1723 CAGATGTGAATGCCCAAAATTAAGCTCGAAGCTTCATCAAACTCCCTCTCGTGT 1782  
Qy 581 ProTyrGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGln 597  
Db 1783 CCTACGGTTCATGGAACCTTCATTCATTCAGGATTCAGGAAACA 1833

RESULT 4  
US-09-758-269-1  
; Sequence 1, Application US/09758269  
; Patent No. US20020104120A1  
; GENERAL INFORMATION:  
; APPLICANT: IUCHI, SATOSHI  
; APPLICANT: KOBAYASHI, MASATOMO  
; APPLICANT: SHINOZAKI, KAZUO  
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN  
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE  
; FILE REFERENCE: 3914-3  
; CURRENT APPLICATION NUMBER: US/09/758, 269  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: JP 2001-003476  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: JP 2000-010056  
; PRIOR FILING DATE: 2000-01-13  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1752  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1749)  
US-09-758-269-1

Alignment Scores:  
Pred. No.: 9,18e-219 Length: 1752  
Score: 1991.00 Matches: 380  
Percent Similarity: 78.44% Conservative: 82  
Best Local Similarity: 64.52% Mismatches: 105  
Query Match: 63.21% Indels: 22  
DB: 10 Gaps: 6

US-09-758-269-6 (1-599) x US-09-758-269-1 (1-1752)  
Qy 16 GlyGlyAsnHisThrGlnProLeuSerSerSerGlnSerSerAspLeuSerTyr--- 34  
Db 31 GGTGGTATTAAACATGGGCT-----CAAGCCCAATGATTTGGGTTTAGG 78  
Qy 35 ---CysSerSerLeuProMetAlaSerArgValThrGlyLeuAsnValSerAla 53  
Db 79 CCCATTAAGAGCAACCGAAGGTTATTAATGACCGTGCAGATCGACGTAACGGA--- 135  
Qy 54 LeuHisThrProAlaLeuHisPheProLysGlnSerSerAsnSerProAlaIleVal 73  
Db 136 TTAACCAAAAACGCCAATATTATTACCCAGAACCCCGCTACTCCGCGC----- 186  
Qy 74 ValLysProLysAlaLysGlnSerAsnThrLysGlnMetAsnLeuPheGlnArgAla 93  
Db 187 -----CAGCAATATCTCTCCGCTAAACATCTTCCAGAAACGCGG\*228

Qy 94 AlaAlaAlaLeuAspAlaAlaGluGlyPheLeuValSerHisGluLysLeuHisProLeu 113  
Db 229 GCATTTGCATTCGACCGGCTGAGCGTGCATTAAATCTCACAGGACAAATCTCCACTT 288  
Qy 114 ProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAlaProValAsnGluGln 133  
Db 289 CCCAAACCGCTGATCCAGTGTTCAGATTGCGGGAATTATTCCTCCGGTACCGGAATCT 348  
Qy 134 ProValArgArgAsnLeuProValValGlyLysLeuProAspSerIleLysGlyValTyr 153  
Db 349 TCCGTCGCGGAAACCTCACCGTCGAAGGAACATCCCTGATGCTATTCACGGTGTTAT 408  
Qy 154 ValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHisHisPhePheAspGly 173  
Db 409 ATCCGTAACCGCGGAATCCGATGTTGAGCCAAACAGCTGGGCACCATTAATTCGACGGA 468  
Qy 174 AspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerTyrAlaCysArgPhe 193  
Db 469 GACGGAATGTTCCAGCAGTTAAATAACACCGGTTACGTAGCTACGATGCCCCGTTT 528  
Qy 194 ThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPheProLysAla 213  
Db 529 AAAAAACCGAGAGATTGGTTTCAGGAAAAACGATGGGTCCGACCATTTTCCGAAAGCA 588  
Qy 214 IleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAla 233  
Db 589 ATCCGCGAGCTTCACGCTCACTCGGAATCGCAGCTTGTATGCTGTTTACGACCGTGG 648  
Qy 234 AlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyr 253  
Db 649 CTITGTGCTGATCAACACCAACACCGCGCTCGAGTAGCAACCGCGTTGGTTTAC 708  
Qy 254 PheAsnGlyArgLeuLeuAlaMetSerGluAspAspLeuProTyrGlnValGlnIleThr 273  
Db 709 TTTAATAACCGGCTTTTAGCTATGTGAGAGACGATTTACCGTACCAATTAATAAATTACT 768  
Qy 274 ProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeuGluSerThr 293  
Db 769 CAACCCGCGATCTCCAAACCGTTGACGTTAGATTCGACGCTCAGTTAAATCGCA 828  
Qy 294 MetIleAlaHisProLysValAspProGluSerGlyGluLeuPheAlaLeuSerTyrAsp 313  
Db 829 ATGATAGCTCACCCGAACCTGGACCGGTTACGAAGGAGCTTCACGCTTAAGCTACGAC 888  
Qy 314 ValValSerLysProTyrLeuLysTyrPheArgPheSerProAspGlyThrLysSerPro 333  
Db 889 GTGTTAAGAAACCTTACCTGAAATTAATTCAGATTTTCGCCAGACGCGGTTAAATCGCCG 948  
Qy 334 AspValGluIleGlnLeuAspGlnProThrMetMetHisAspPheAlaIleThrGluAsn 353  
Db 949 GAATGGAGATCCCGCTCGAAACTCCGACGATGATTTCAGATTCGCTATAACGAGAAAT 1008  
Qy 354 PheValValValProAspGlnGlnValValPheLysLeuProGluMetIleArgGlyGly 373  
Db 1009 TTTGTGTGATTTCTGATCAACAGTCGTGTTCAAGCTCGCGAGATGATTTCCGGTAA 1068  
Qy 374 SerProValValTyrAspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyrAla 393  
Db 1069 TCTCCGTTGTTTCGACGCGAGAAAGGTTTCCCGATTGGGGATAATGCCAAGGACGCG 1128  
Qy 394 GluAspSerSerAsnIleLysTyrIleAspAlaProAspCysPheCysPheHisLeuTyr 413  
Db 1129 ACAGAAGCTTCTCAGATAATCTGGGTGAATCTCCGAGACGCTGCTGTTTCTATCTCG 1188  
Qy 414 AsnAlaTyrGluProGluThrAspGluValValIleGlySerCysMetThrPro 433  
Db 1189 AATGATGGGAATCCCGGAGACGAGGAGATTTGGTGATCGGATCGTGTATGTCGCGC 1248  
Qy 434 ProAspSerIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArg 453  
Db 1249 GCGGATTCATCTTCAACGAGAGACGAGAGCTTGAAGACGCTTTTGTCCGAGATCAGG 1308  
Qy 454 LeuAsnLeuLysThrGlyGluSerThrArgProIleIleSerAsnGluAspGlnGln 473



1309	DB	ATAAACCTCAGAACACGTTAAACACCCGCGTCGTTGGTTGGTTAACGAGGAT-----	1362
474	QY	ValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPheAla	493
1363	DB	GTAATATTAGAGATTGGTATGGTTAACCGGAACCGGTTAGAGAAACCCGGTTCGCG	1422
494	QY	TyrLeuAlaLeuAlaGluProTprProLysValSerGlyPheAlaLysValAspLeuThr	513
1423	DB	TTTTTGGCTATTGCTTATCCTTGGCCAAAAGTTTCGGTTTCGCTAAGGTGATCTTGC	1482
514	QY	ThrGlyGluValLysLysHisLeuTyrGlyYAspAsnArgTyrGlyGlyGluProLeuPhe	533
1483	DB	ACCGGTGAGATGAAAAAATATATTATCGCGCGGTGAGAAATATGCGCGCAACCGTTTTTC	1542
534	QY	LeuProGlyGluGlyGly-----GluGluAspGluGlyTyrLleLeuCysPhe	549
1543	DB	TTCCCGCGCAACTCCCGGTAAACGCGAAGAAAATGAAGATACCGGTTATATTTTGTCAAC	1602
550	QY	ValHisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaValSerLeuGlu	569
1603	DB	GTTCATGACGAAGAACCAAGACATCAGAGCTTCAGATTATTAAACGTGTGTTAATTAAAG	1662
570	QY	ValGluAlaThrValLysLeuProSerArgValProTyrGlyPheHisGlyThrPheIle	589
1663	DB	CTTGAACCTACGATTAAACATACCGCTCAGAGTACCGTATCGGTTTCATGSCACATTGTG	1722
590	QY	GlyAlaAspAspLeuAlaLysGlnVal	598
1723	DB	GATTCCGAATGAACCTGGTATCAATTA	1749

## RESULT 5

```

US-09-758-269-13
; Sequence 13, Application US/09758269
; Patent No. US2002010120A1
; GENERAL INFORMATION:
; APPLICANT: UUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO.13

```

Alignment Scores:		
Pred. No.:	1.03e-211	Length:
Score:	1930.00	Matches:
Percent Similarity:	75.68%	Conservative:
Best Local Similarity:	63.34%	Mismatches:
Query Match:	61.37%	Indels:
DB:	10	Gaps:
		1815
		385
		73
		122
		22
		8

US-09-758-269-6 (1-599) x US-09-758-269-13 (1-1815)

Qy	22	Pro	Leu	Ser	Ser	Ser	Gln	Ser	Ser	Asp	Leu	Ser	Tyr	Cys	Ser	Ser	Leu	Pro	Met	Ala	41
Db	49	CGG	CCC	CGG	TCC	AGG	CCG	CGG	GCT	TCC	-----	AA	TCC	GTC	AGG	TTC	CTC	G	93		

42 SerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrPro-----Pro 58  
 94 CCGCGCGCGTGCAGTCCGTCGCGCGCGCGAGTGCCTCCAGGCGCGTTCACAAAGGCC 153  
 59 AlaLeuHisPheProLysGlnSerSerAsnSerProAlaIleVal----- 74  
 154 GTCGCGCAGCTGCTCGCGCGTCCAGGAAGCGCGCGCATTCGCGTCCAGGGCAGCGC 213  
 75 ---LysProLysAlaLysGlnSerAsnThrLysGlnMetAsnLeuPheGlnArgAla 93  
 214 CGCGCGCGCAGGAAACGGAGGGCGGCAAGAGCAGCTCACTTGTTCAGCGCGCGCGG 273  
 94 AlaAlaLeuAspAla---AlaGluGlyPheLeuValSer---HisGluLysLeuHis 111  
 274 CGCGCGCGCGCTGCACGCGTTCAGGAAGAGGTTCGTGGCAACGCTCTCGAGCGGCGCCAC 333  
 112 ProLeuProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAlaProValAsn 131  
 334 GCGCTGCCCAGCACGCGCGACCGCGCGTGCAGATCGCGGCAACTTCGGGCGCGTCGGG 393  
 132 GluGlnProValArgAsnLeuProValValGlyLysLeuProAspSerIleLysGly 151  
 394 GAGAGCGCGCGCTGCACGAGTCCCGCTCCGCGCGCATCCCGCTTCATCAGCAGGG 453  
 152 ValTyrValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHisPhePhe 171  
 454 GTCTAGCGCGCAACGCGCGCAACCTCTCTCGACCCCGTTCGGGGGACCACTCTTC 513  
 172 AspGlyAspGlyMetValHisAlaValLysPheGluHisGlySerAla---SerTyrAla 190  
 514 GAGCGGACGCGCATGTCACGCGCTCGCGATACACGACGCGCGCGCGAGTCTACGCGC 573  
 191 CysArgPheThrGlnThrAsnArgPheValGlnGluArgGluLeuGlyArgProValPhe 210  
 574 TGCGCTTCACGAGACGCGCGCTCGCGCAGGAGCGCGCATCGCGCGCGCTCTTC 633  
 211 ProLysAlaIleGlyGlnLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyr 230  
 634 CCCAAGGCCATTGCGAGTGCACGGGCACTCGGATCGCGCGCTCGCCCTGTCTCTAC 693  
 231 AlaArgAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsnAlaGly 250  
 694 GCGCGCGCGCTGCGCGCTCTGTGGACCTCGCGCGCACCGCGTGGCCAAACGCGGC 753  
 251 LeuValTyrPheAsnGlyArgLeuLeuAlaMetSerGluAspLeuProTyrGlnVal 270  
 754 CTCGCTACTTCAACGCGCGCTGCTCGCATGTCGAGAGCAGCTCCCTTACCACGCTC 813  
 271 GlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeu 290  
 814 CGCGTGGCGACGACGCGGACCTCGACCGTGGCGCGTACGACTTCACGGGAGCTC 873  
 291 GluSerThrMetIleAlaHisProLysValAspProGluSerGlyGlnLeuPheAlaLeu 310  
 874 GGCTGGCGCCATGATCGCGCACCCCAAGCTGACCCCGGACCGGGAGTTCACGCGCTC 933  
 311 SerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerProAspGlyThr 330  
 934 AGCTACGACGTCATCAAGAGCGCGTACCTCAAGTACTTCTACTTCAGGCGCGACGACC 993  
 331 LysSerProAspValGluIleGlnLeuAspGlnProThrMetMetHisAspPheAlaIle 350  
 994 AAGTCCGACGACGCTGGAGATCCCGCTGGAGCAGCCCGCATGATCCACGACTTCGCGCATC 1053  
 351 ThrGluAsnPheValValProAspGlnGlnValValPheLysLeuProGluMetIle 370  
 1054 ACCGAGAACTTCGTTGGTTGGCCCGACCACTGCTGTTCAAGTCTCCAGGAGATGCTG 1113  
 371 ArgGlyGlySerProValTyrAspLysAsnLysValAlaArgPheGlyIleLeuAsp 390  
 1114 CGCGCGCGGTCCCGGTGGTCTGGACAAAGAGAAGACGTCGCGGTTCGCGGTCTCCCC 1173  
 391 LysTyrAlaGluAspSerSerAsnIleLysTrpIleAspAlaProAspCysPheCysPhe 410

```

Db 1174 AAGCAGCGCGGAGCGCTCGAGATGGCGTGGGTGGAGCGCGGAGCTGCTTCTGCTTC 1233
QY 411 HisLeuTrpAsnAlaTrpGluGluProGluThrAspGluValValValIleGlySerCys 430
Db 1234 CACCTGTGGACAGCGGTGGAGAGCGAGCGGCGGAGGTGGTGTGATCGGCTCTGCG 1293
QY 431 MetThrProProAspSerIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSer 450
Db 1294 ATGACCCCGCGGAGCTCCATCTCAACAGATCCAGCAGCGCTGGAGAGCGTCTGAC 1353
QY 451 GluIleArgLeuAsnLeuLysThrGlyGluSerThrArgArgProlIleSerAsnGlu 470
Db 1354 GAGATCCCGCTGGAGCGCGGAGCGGCGGCTCCAGCGCGCGCGCTCCCTGCGCG --CCG 1410
QY 471 AspGlnGlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThr 490
Db 1411 TCGCAGCAGGAGAACCTGAGGTGGCATGCTGAACCGAACCTGCTGGGCGGAGAGC 1470
QY 491 LysPheAlaTrpLeuAlaLeuAlaGluProTrpProLysValSerGlyPheAlaLysVal 510
Db 1471 CGGTACGGGTACCTCGCGGTGGGAGCGCGGTGGCCCAAGGAGTGGGCTTCGCCAAGGAG 1530
QY 511 AspLeuThrThrGlyGluValLysLysHisLeuTrpGlyAspAsnArgTrpGlyGlyGlu 530
Db 1531 GACCTGTCCAGCGGAGCTCACCAAGTTCAGATACGGCGGAGGCGCGGCTTCGGCGGCGAG 1590
QY 531 ProLeuPheLeuProGlyGlyGlyGlu-----GluAspGluGlyTyr 545
Db 1591 CCCTGCTCGTTCCTCCATGACCGCGCGCGCGCCGCCACCGCGCGCGGAGCAGCGGTAC 1650
QY 546 IleLeuCysPheValHisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAla 565
Db 1651 GTGCTCACTTCGTCACAGCAGAGCGCGCGCGCGCGCGCTCGAGCTACTGTGTCAATGCC 1710
QY 566 ValSerLeuGluValGluAlaThrValLysLeuProSerArgValProTrpGlyPheHis 585
Db 1711 GCGCAGATCCGCGCTCGAGGCGCAGCGTTCAGTGGCGTCCCGCGTGGCGCTTCGCTTCCAC 1770
QY 586 GlyThrPheIleGlyAlaAspAspLeuAlaLysGln 597
Db 1771 GGCACCTTCATCAGGGCGAGGAGCTCGAGGCCGAG 1806

```

## RESULT 6

```

US-09-758-269-9
; Sequence 9, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: IUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 1734
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1731)
US-09-758-269-9

```

Alignment Scores:

Pred. No.: 4,836-181 Length: 1734

```

Score: 1663.50 Matches: 323
Percent Similarity: 72.56% Conservative: 100
Best Local Similarity: 55.40% Mismatches: 137
Query Match: 52.81% Indels: 23
DB: 10 Gaps: 9

US-09-758-269-6 (1-599) x US-09-758-269-9 (1-1734)

QY 24 LeuSerSerSerGlnSerSerAspLeuSerTrpCysSerSerLeuProMetAlaSerArg 43
Db 25 CTTCTCCGACGAGACTTCTCTCTGTTCTATTCTTCCACAAACCCAAAAATGCAAT 84
QY 44 ValThrArgLysLeuAsnValSerSer---AlaLeuHisThrProProAlaLeuHisPhe 62
Db 85 ATTCTCGACGAAATCTCTATTAAACCTTTCAGATCCCTTCAAGATCCGACACTTCTCTACTTCT 144
QY 63 ProLysGlnSerSerAsnSerProAlaIleValLysProLysProLysAlaLysGluSerAsn 82
Db 145 CCGGTTCCGTCACCG-----GTTAAGCTCAAACCAACGATCCAAAC 186
QY 83 ThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAlaGlu--- 101
Db 187 -----TTAAACCTTCTTCAAGAGCTAGCGGTACGATGCTCGCAAGATTGAGTCC 237
QY 102 GlyPheLeuValSerHisGlyLysLeuHisProLeuProLysThrAlaAspProSerVal 121
Db 238 TCTATCGTTATTCCTATGGAGCAGATCGCCGCTTCTTAACCCGACCGCGCGGTT 297
QY 122 GlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuProVal 141
Db 298 CAATTATCAGTAACCTTCGCTCCGTTAATGAATGTCGCGTTTCAAGCGGTTTGAAGTG 357
QY 142 ValGlyLysLeuProAspSerIleLysGlyValTrpValArgAsnGlyValAsnProLeu 161
Db 358 GTTGTGCAATTCCTTCTTGTCTAAAGAGTTTACATCCGTAAACGGTGCAACCCCTATG 417
QY 162 HisGluProValTrpGlyHisHisPheAspGlyAspGlyMetValHisAlaValLys 181
Db 418 TTTCCGCGCTTAGCGCGGACATCATTTATTGACGGTACGGAATGATTCAACCGCGTTAGT 477
QY 182 -----PheGluHisGlySerAlaSerTrpAlaCysArgPheThrGlnThrAsnArgPhe 199
Db 478 ATCGGTTTTGTATAAC---CAGGTTAGTTACAGCTCCGCTACACTAAACAAACCGGCTT 534
QY 200 ValGlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyLeuHisGly 219
Db 535 GTTCAAGAAACCGCGCTTCGACGATCGGTTTTTCCCTAAACCAATCGGCGAGCTTCACGCG 594
QY 220 HisThrGlyIleAlaArgLeuMetLeuPheTrpAlaArgAlaAlaAlaGlyIleValAsp 239
Db 595 CATTCGCGTCTAGCTCGACTCGCTCTCTTCACGGCTCGAGCTGGGATCGGTCTAGTGGAC 654
QY 240 ProAlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTrpPheAsnGlyArgLeuLeu 259
Db 655 GGGACACGTGGCATGGCGGTAGCTAACGCGGTGTGGTTTTCTTTAAACGGCAGGTTATTA 714
QY 260 AlaMetSerGluAspAspLeuProTrpGlnValGlnIleThrProAsnGlyAspLeuLys 279
Db 715 GCCATGTCAGAAAGATGATCTTCTTACCAGGTGAAGATCGACGGTCAAGGAGATCTTGAG 774
QY 280 ThrValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLys 299
Db 775 ACGATCGGACGGTTCGATTTCGATGACCGAGATTGACTCTTCAGTGATAGCGGATCTTAAG 834
QY 300 ValAspProGluSerGlyGluLeuPheAlaLeuSerTrpAspValValSerLysProTrp 319
Db 835 GTGGACCGGACACAGGAGATCTCCATACACTGAGCTACACGTTTGAAGAAACCTCAT 894
QY 320 LeuLysTrpPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeu 339
Db 895 CTCAGGTATCTTAATTTCAACACGTCGCGGAAAAGACACGTCGACGTGGAGATCAGCTTC 954
QY 340 AspGlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAsp 359

```





629	DB	CGCGCTCACTAGCTCGTGGAGCTTTAAACGGCAGCTAGGCTTTTACC	CGCAGAGTAAATC	688
240	QY	roAlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuA	260	
689	DB	CGGTAAACGGCATTTGGTTTAGCTAATACAAAGTCTAGCTTCTTCAGTAAACCGTCTCTTTG	748	
260	QY	laMetSerGluAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysT	280	
749	DB	CTTTAGGTGAACTGATTTTACCCTACCGCTCCGATTACCAATCAGGAGATATTGAA	808	
280	QY	hrValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysV	300	
809	DB	CGATCGGACGGTACGATTTTCGACGGGAATATACGATGAGTATGACAGCTCATCTCAAA	868	
300	QY	alAspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrL	320	
869	DB	CCGATCCAATAACCGGAGAACTTTCGCTTCGGTACGGTCCGGT---CCACCGTTT	925	
320	QY	eulysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluIle---	339	
926	DB	TAACATATTTCCCGTTTGATTCGCCCGGAAAAACAAAGACAGCTTCGATATTCTCGA	985	
339	QY	euAspGlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProA	359	
986	DB	TGACGCTCCGTCGTTTCTCCATGACTTCGCGATCACGAAACGTCACGCGATTTTCGCAG	1045	
359	QY	spGlnGlnValValPheLys-----LeuProGluMetIleArgGlyGlySerProV	376	
1046	DB	AGATTACGCTTGGCATGAGGATGACATGTGGATTGGTTCTTCGAAGTGGTTCTCCGG	1105	
376	QY	alValTyrAspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspS	396	
1106	DB	TTGGTACTGATAACGGAAAAACCTCAAGGCTTGGAGTGATTCCTTAAGTACCGCGGAGATG	1165	
396	QY	erSerAsnIleLysTrpIleAspAlaProAspCysPheCysPheHisLeuTyrAsnAlaIat	416	
1166	DB	AGTCGGAGATGAATGGTTCGAAGTTCCTGGATTCAATATCATTCACGCTATTAAATGCTT	1225	
416	QY	rpGluGluProGluThrAspGluValValIleGlySerCysMetThrProProAspS	436	
1226	DB	GGGATGAAGATGATGTAACAAACAGCGCTCGTTTGTATGTCACCGCAATATTATGTCGATTTGAAC	1285	
436	QY	erIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnL	456	
1286	DB	ATACTTTAGAGAGATGAT---CTGGTTCATGCTTTGGTGGAGAAAGTGNAGATCGATC	1342	
456	QY	euLysThrGlyGluSerThrArgArgProIleIleSerAsnGluAspGlnValAsnL	476	
1343	DB	TCGTCACCGGATTTGTGAGAGCTCATCGATCTCAGCGAGG-----AATC	1387	
476	QY	euGluAlaGlyMetValAsnArgAsnMetLeuGlyValGlyLysThrLysPheAlaTyrLeuA	496	
1388	DB	TCGATTCGCTGTGATTAATCCGGCGTTTCTCCGGAGATGTAGCAGGTACGTTTACCGGG	1447	
496	QY	laLeuAlaGluProTrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyG	516	
1448	DB	CGATTTGGAGATCCGATCGCAAGATCTCCGCTGTGTGTAAGACTTGTGTCTAAAGGAG	1507	
516	QY	lu-----VallysLysHisLeuTyrGlyAspAsnArgTyrGlyGlyGluP	531	
1508	DB	ATCGGGATGATTGTACGGTGGCCGTAGATGTACGGTTCAGTTGTTTACCGCGGAGAAC	1567	
531	QY	roLeuPheLeuProGlyGluGlyGly-----GluGluAspGluGlyTyrIleL	547	
1568	DB	CGTTTTCGTAGCTAGGATCTCGTAAATCCGGAGCGGAGGAGATGATGTTATGTGG	1627	
547	QY	euCysPheValHisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaValS	567	
1628	DB	TGACGTATGTTTACGATGAAGTACTCGAGATCCGAAGTTTCTGCTGTGATGACGCTAAAT	1687	
567	QY	er-----LeuGluValGluAlaThrValLysLeuProSerArgValProTyrGlyPheH	585	
1688	DB	CGCCGAGGTTTGAATTCGTCGCCCGCGTGAAGTTCGCCGAAGGTTTCGTCACGATTC	1747	

```

596 585 isg1ythrPharilec1yalaasprleueu1a1y5 596
597 1748 ATGGGTTATTGTTCAAGGAAGTGCACCTTAATAG 1782
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
```

[illegible]

Db 382 GAAGAGTTCTTCGAGAGTCCCAATTCATGAAG---ATTGGTGACCTTAAGGGGTTTTC 438  
Qy 222 GlytlealaArgLeuMetLeuPheTy:AlaArgAlaAlaAlaGlyIleValAspProAla 241  
Db 439 GGAATGTAAATGGTCAATCTCAACAGCTGAGAACGAAGCTCAAAATATTGGCAACACT 498  
Qy 242 HisGlyThrGlyValAlaAsnAlaGlyLeuValTy:PheAsnGlyArgLeuLeuAlaMet 261  
Db 499 TATGGAATGGAATGCCAATACAGACTCTGTATATCACCATTGGAATACTTATAGCATTA 558  
Qy 262 SerGluAspLeuProTy:GlnValGlnIleThrProAsnGlyAspLeuLysThrVal 281  
Db 559 CAGGAGGCAGATAAGCGTACGTCAATCAAGATTGGAGATGAGACCTGCCAACTCTT 618  
Qy 282 GlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysValAsp 301  
Db 619 GGTATAATAGATTATGACAGAGATTGACCCACTCTCTCACTGCTCAACCCAAAGTTGAC 678  
Qy 302 ProGluSerGlyGluLeuPheAlaLeuSerTy:AspValValSerLysProTy:LeuLys 321  
Db 679 CCGTTACGGGTGAATGTTTACATTCCGCTATTTCG---CATAGCCACCTTATCTCACA 735  
Qy 322 Ty:PheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAspGln 341  
Db 736 TACAGAGTTATCTCGAAGATGGCATTATGCATGACCCAGTCCCAATTACTATATCAGAG 795  
Qy 342 ProThrMetHisAspPheAlaIleThrGluAsnPheValValProAspGlnGln 361  
Db 796 CCTATCATGATGATGATTGCTATTACTGAGACTTATGCAATCTTCATGATCTTCCT 855  
Qy 362 ValValPheLysLeuProGluMetIleArgGlyGlySerProValVal---Ty:AspLys 380  
Db 856 ATGCACCTTCAGCCCAAGGAATGGTGAAGAGAAGAAATGATATATATCATCTATGATCC 915  
Qy 381 AsnLysValAlaA:ArgPheGlyIleLeuAspLysTy:AlaGluAspSerSerAsnIleLys 400  
Db 916 ACAAAGAGCTCGTTTGGTGTCTCCAGCTATGCCAGGATGAATGATGATTAGA 975  
Qy 401 TrpIleAspAlaProAspCysPheHisLeuTrpAsnAlaTrpGluGluProGlu 420  
Db 976 TGGTTGAGCTCCCACTGCTTTATTTCCACAACGCCAATGCTTGGAA---GAA 1029  
Qy 421 ThrAspGluValValIleGlySerCysMetThrProProAsp----- 435  
Db 1030 GAGGATGAAGTCGCTCTCATCTTCTGCTGCTGAGAAATCCAGATCTTGACATGGTCAGT 1089  
Qy 436 SerIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsn 455  
Db 1090 GGGAAAGTGAAGAAAGAACTCGAAATTTGGCAACGAACTGTACGAAATGAGATTCAAC 1149  
Qy 456 LeuLysThrGlyGluSerThrArgArgProIleIleSerAsnGluAspGlnValAsn 475  
Db 1150 ATGAAACGGCTCAGCTCTCTCAAAAAAATATCCGATCTGCG----- 1194  
Qy 476 LeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTy:Leu 495  
Db 1195 GTTGATTTCCCAAGATCAATGAGTGCTPACCCGGAAAGAAACAGATACGTATATGGA 1254  
Qy 496 AlaLeuAlaGluProTrpProLysValSerGlyPheAlaLysValAsp----- 511  
Db 1255 ACAATTCTGGACAGTATCGAAAGGTTTACCGGAATCATCAAGTTTGTATCTGCATGCAGAA 1314  
Qy 512 -----LeuThrThrGlyGluValLysLysHisLeuTy:----- 522  
Db 1315 CTGAGACAGGGAAGAAATGCTGGAAGTACGAGGTAATATCAAGGAATATATGACCTG 1374  
Qy 523 GlyAspAsnArgTy:GlyGlyGluProLeuPheLeuProGlyGluGlyGlyGluGluAsp 542  
Db 1375 GGAGAAGCAGATATGTTTCAGAGCTATCTATGTTCCGCGTGAGACAGCAGAGAAGAC 1434  
Qy 543 GluGlyTy:LeuCysPheValHisAspGluLysThrTrpLysSerGluLeuGlnIle 562  
Db 1435 CACGCTTACTGATATCTTTGTTTCATGATGAACACAGCGGAATATCATCGTACCTGTG 1494

Qy 563 ValAsnAlaValSerLeuGluValGlu-----AlaThrValLysLeuProSerArgVal 580  
Db 1495 ATAGACGCAAAACAAATCGCTCGCTGAACCGGTGGAGTGGTGGAGCTCCGACAGGTC 1554  
Qy 581 ProTy:GlyPheHisGlyThr:PheIleGlyAlaAspLeuAlaLysGlnValVal 599  
Db 1555 CCATATGCTTCCTCCATGCTTGTGTTTACAGAGGACAACTCCAGGAACAACCTCTT 1611

## RESULT 10

US-09-758-269-17  
; Sequence 17, Application US/09758269  
; Patent No. US20020104120A1  
; GENERAL INFORMATION:  
; APPLICANT: UCHI, SATOSHI  
; APPLICANT: KOBAYASHI, MASATOMO  
; APPLICANT: SHINOZAKI, KAZUO  
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN  
; FILE REFERENCE: 3914-3  
; CURRENT APPLICATION NUMBER: US/09/758,269  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: JP 2001-003476  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: JP 2000-010056  
; PRIOR FILING DATE: 2000-01-13  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 17  
; LENGTH: 1617  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1614)  
US-09-758-269-17

## Alignment Scores:

Pred. No.:	2,148-97	Length:	1617
Score:	937.00	Matches:	207
Percent Similarity:	55.81%	Conservative:	105
Best Local Similarity:	37.03%	Mismatches:	199
Query Match:	29.75%	Indels:	48
DB:	10	Gaps:	13

US-09-758-269-6 (1-599) x US-09-758-269-17 (1-1617)

Qy 64 LysGlnSerSerAsnSerProAlaIleValLysProLys---AlaLysGluSerAsn 82  
Db 10 AAACCTCAGTGATGGCAGCATCATCTCAGTCCATCTAGACCCCTCCAGAGGTTTCTCC 69  
Qy 83 ThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAlaAlaGluGly 102  
Db 70 TCGAAGCTTCTCGATCTTCTCGAGAGACTGTGTCGAAGCTCATG----- 114  
Qy 103 PheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSerValGln 122  
Db 115 -----CACGATGCTTCTCTCCCTCTCCAC-----TAC 141  
Qy 123 IleAlaGlyAsnPheAlaProVal---AsnGluGlnProValArgArgAsnLeuProVal 141  
Db 142 CTCTCAGGCAACTCTCGCTCCCATCCGTGATGAAATCTCTCCGTCAGAGATCTCCCGCTC 201  
Qy 142 ValGlyLysLeuProAspSerIleLysGlyValTy:ValArgAsnGlyAlaAsnProLeu 161  
Db 202 CATGGATTTCTTCCCGAATGCTTGAATGGTGAATTTGAGGTTGGTCCAAACCCCAAG 261  
Qy 162 HisGluProValThrGlyHisHisPhePheAspGlyAspGlyMetValHisAlaValLys 181  
Db 262 TTTGATGCTGTGCTGGATATCATCTGTTGATGAGATGGGATGATTCATGGGTACGC 321  
Qy 182 PheGluHisGlySerAlaSerTy:AlaCysArgPheThrGlnThrAsnArgPheValGln 201

Db 322 ATCAAGAGTGGAAAGCTACTTATGTTCTCGATATGTTAAGACATCATCGTCTTAAACGAG 381  
Qy 202 GluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHisThr 221  
Db 392 GAAGAGTCTTCGGAGCTGCCAAATTCATGAG---ATTGGTGACCTTAAGGGGTTTTC 438  
Qy 222 GlyIleAlaArgLeuMetLeuPheTyraAlaArgAlaAlaGlyIleValAspProAla 241  
Db 439 GGATTGCTTAATGGTCAATATCAACAGCTGAGAACGAAAGCTCAAAATATTGGCAACACT 498  
Qy 242 HisGlyThrGlyValAlaAsnAlaGlyLeuValTyraPheAsnGlyArgLeuLeuAlaMet 261  
Db 499 TATGGAATGGAACTGCCAATACACCACTCGTATATACCATCGGAAACACTTACGATTA 558  
Qy 262 SerGluAspLeuProTyroGlnValGlnIleThrProAsnGlyAspLeuLysThrVal 281  
Db 559 CAGGAGGCGCATTAAGCGGTACCTCAAAAGTTTGGAGATGGAGAGCTCCCAACCTCT 618  
Qy 282 GlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysValAsp 301  
Db 619 GGTATAATAGATTATGACAGAGATTGCCCACTCTTCACTCTGCTCCCAACCAAGTTGAC 678  
Qy 302 ProGluSerGlyGluLeuPheAlaLeuSerTyraPheValValSerLysProTyroLys 321  
Db 679 CCGGTTACGGTGAAATGTTTACATTCGGCTATTGCG---CATACGCCACCTTATCTCACA 735  
Qy 322 TyraPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAspGln 341  
Db 736 TACAGATTATCTGAAAGATGCGATTATGATGATGCCAGCTCCCAATTAATATATACAG 795  
Qy 342 ProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAspGlnGln 361  
Db 796 CCTATCATGATGATGATTTGCTATTACTGAGACTTATGCAATCTTCATGATCTTCCT 855  
Qy 362 ValValPheLysLeuProGluMetIleArgGlyGlySerProValVal---TyraPlys 380  
Db 856 ATGCACCTCAGGCCAAAGAAATGGTGAAGAGAGAAATGATATATCTCATTTGATCCC 915  
Qy 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyraAlaGluAspSerSerAsnLys 400  
Db 916 ACAAAGAGCGTCGTTTGGTGTCTTCGCGCTATGCCAGATGAACCTTATGATTAGA 975  
Qy 401 TrpIleAspAlaProAspCysPheCysPheHisLeuTrpAsnAlaTrpGluProGlu 420  
Db 976 TGGTTTGGAGCTTCCCACTGCTTTATTTCCCAACAGCCAAATGCTTGGAA-----GAA 1029  
Qy 421 ThrAspGluValValIleGlySerCysMetThrProProAsp----- 435  
Db 1030 GAGATGAAGTCTCTCATCTACTTCTGCTTGGATCCAGATCTTGACATGTCAGT 1089  
Qy 436 SerIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsn 455  
Db 1090 GGGAAAGTGAAGAAAGAACTCGAAATTTTGGCAACAGAACTGTACGAATGAGATTCAAC 1149  
Qy 456 LeuLysThrGlyGluSerThrArgArgProIleIleSerAsnGluAspGlnGlnValAsn 475  
Db 1150 ATGAAACGGGCTCAGCTTCTCAAAAAAACHATCCGATCTGCG----- 1194  
Qy 476 LeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyroLeu 495  
Db 1195 GTTGATTTCGCCAGATCAATAGTGTCTCACCGGAAAGAAACAGAGATACGTATATGGA 1254  
Qy 496 AlaLeuAlaGluProTrpProLysValSerGlyPheAlaLysValAsp----- 511  
Db 1255 ACAATTCTGACAGTATCGCAAGGTTACCGGATCATCAAGTTTGTATCTGATCGATCGA 1314  
Qy 512 -----LeuThrThrGlyGluValLysHisLeuTyro----- 522  
Db 1315 GCTGAGACAGGAAAGAAATGCTGGAGTAGGAGGTAATATCAAGGAATATATGACCTG 1374  
Qy 523 GlyAspAsnArgTyroGlyGluProLeuPheLeuProGlyGlyGlyGluGluAsp 542  
Db 1375 GGAAAGGCGAGATATGCTCAGAGGCTATCTATGTTCCCGGTGAGACAGCAGAGAGAC 1434

Qy 543 GluGlyTyroIleLeuCysPheValHisAspGluLysThrTrpLysSerGluLeuGlnIle 562  
Db 1435 GACGGTTACTTGTATATTCTTTGTTGATGATGAAACACAGGGAAATCATGGTGTACTGTG 1494  
Qy 563 ValAsnAlaValSerLeuGluValGlu-----AlaThrValLysLeuProSerArgVal 580  
Db 1495 ATAGAGCGCAAAACATATGTCGCTGACCGGTGGAGTGTGGAGCTCCGCGACAGGTC 1554  
Qy 581 ProTyroGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGlnValVal 599  
Db 1555 CCATATGGTTCATGCTTGTGTTGTTACAGAGGAACCACTCCAGGAACAAACTCTT 1611

RESULT 11  
US-09-878-574-2543  
; Sequence 2543, Application US/09878574  
; Patent No. US20020110548A1  
; GENERAL INFORMATION:  
; APPLICANT: Byrum, Joseph R.  
; APPLICANT: Thompson, Michael D.  
; APPLICANT: La Rosa, Thomas J.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(15401)B  
; CURRENT APPLICATION NUMBER: US/09/878,574  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/333,535  
; PRIOR FILING DATE: 1999-06-14  
; NUMBER OF SEQ ID NOS: 15775  
; SEQ ID NO 2543  
; LENGTH: 393  
; TYPE: DNA  
; ORGANISM: Glycine max  
; OTHER INFORMATION: Clone ID: L1B3028-023-Q1-B1-G10  
US-09-878-574-2543

Alignment Scores:  
Pred. No.: 158-50 Length: 393  
Score: 522.00 Matches: 95  
Percent Similarity: 85.16% Conservative: 14  
Best Local Similarity: 74.22% Mismatches: 19  
Query Match: 16.57% Indels: 0  
Gaps: 0

US-09-758-269-6 (1-599) x US-09-878-574-2543 (1-393)

Qy 328 AspGlyThrLysSerProAspValGluIleGlnLeuAspGlnProThrMetHisAsp 347  
Db 8 GACGAGAGGAAGTCGCGGACATAGAAATTCCTTTGACGCGCGGACGATGACGACGAC 67  
Qy 348 PheAlaIleThrGluAsnPheValValProAspGlnGlnValValPheLysLeuPro 367  
Db 68 TTCGCAATCAGGAGAAATTCGTTGGTGTATCCCGGACACAGAGTGTGTTCAGCTCGC 127  
Qy 368 GluMetIleArgGlyGlySerProValValTyraPheAsnLysValAlaArgPheGly 397  
Db 128 GAAATGATCAAGAGGATCGCGGTGATCTACGCGCGGAGAAATTCGCGGTTTCGA 187  
Qy 388 IleLeuAspLysTyroAlaGluAspSerSerAsnIleLysTrpIleAspAlaProAspCys 407  
Db 188 ATATGCCAAGTACGCTTCGCGAGCGTCCAGCATCTGTGGTGGTACTCGCCAGACAGC 247  
Qy 408 PheCysPheHisLeuTrpAsnAlaIleTrpGluProGluThrAspGluValValIle 427  
Db 248 TTCCTCTTCACCTCTTGGAAACGCTGGAGGAACCGCAACAAAGTGTGTGTAATA 307  
Qy 428 GlySerCysMetThrProProAspSerIlePheAsnGluSerAspGluAsnLeuLysSer 447  
Db 308 GGTTCGTGATGACACCCAGATTCATATTCATATGACAGAGAGAGACTGTAAGAC 367  
Qy 448 ValLeuSerGluIleArgLeuAsn 455  
Db 368 GTTTTACAGAGTAAGCTGAAC 391



RESULT 12  
US-08-976-063C-1/c  
Sequence 1, Application US/08976063C  
Publication No. US20020182697A1  
GENERAL INFORMATION:  
APPLICANT: Alexander Steinbuechel; Horst Priefert; Jurgen Rabenhorst  
TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF  
TITLE OF INVENTION: CONFERYL ALCOHOL, CONFERYLALDEHYDE, FERULIC ACID, VANILLIN A  
TITLE OF INVENTION: ACID AND THEIR USE  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SPRUNG KRAMER SCHAEFER & BRISCOE  
STREET: 660 White Plains Road  
CITY: Tarrytown  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10591-5144  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage  
COMPUTER: HP VECTRA  
OPERATING SYSTEM: DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/976.063C  
FILING DATE: 21-NOV-1997  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 196 49 655.1 (Germany)  
FILING DATE: 29-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Kurt G. Briscoe  
REGISTRATION NUMBER: 33,141  
REFERENCE/DOCKET NUMBER: Bayer 9998-CAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (914) 332-1700  
TELEFAX: (914) 332-1844  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 32679 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (Genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Pseudomonas sp.  
STRAIN: HR199  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3146..3997  
OTHER INFORMATION: /gene= "ORF1"  
US-08-976-063C-1

Alignment Scores:  
Pred. No.: 1,128-35 Length: 32679  
Score: 419.00 Matches: 160  
Percent Similarity: 37.97% Conservative: 83  
Best Local Similarity: 25.00% Mismatches: 229  
Query Match: 13.30% Indels: 168  
DB: 8 Gaps: 24

US-09-758-269-6 (1-599) x US-08-976-063C-1 (1-32679)

QY 19 HisThrGlnProLeuSerSerGlnSerSerAspLeuSerTyrCysSerSerLeu 38  
DB 17011 CATACGGCGCCACCAAGTAGTGTTCATCTTGAATTCCTAAACATAGCCGCGCTCGGC 16952  
QY 39 ProMetAlaSerArgValThrArgLysLeuValSerAlaLeuHisThrProPro 58  
DB 16951 CCATTTTCA-----ATACCACCA 16934

QY 59 AlaLeuHisPheProLysGlnSerSerAsn-----SerProAlaIleValVal 74  
DB 16933 CCATTGCACCTCACAGACATCTTCACTTCACCGGCACACATGAAATCCG-----GTC 16883  
QY 75 LysProLysAlaLysGlnSerAsnThrLys-----GlnMetAsnLeu 88  
DB 16882 AGTCCTAAACATAACTCCACCGCGGACCGCGCATTTAGGCCATTAGCCATTCAGAACACAAA 16823  
QY 89 PheGlnArgAlaAlaAlaAlaAlaLeuAspAlaAlaGluGlyPheLeuValSerHisGlu 108  
DB 16822 GGAGACCGTGCATCGCAGATTC----- 16799  
QY 109 LysLeuHisProLeuProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAla 128  
DB 16798 -----AACGGCAACGACCG-----CAATTAGTAGAGACACATTCCTC 16763  
QY 129 ProValAsnGluGlnProValArgAsnLeuProValGlyLysLeuProAspSer 148  
DB 16762 CCCACCGGTATAGAGCGACACTTGTTCGATCTAGAGGTGACGGCGAAATCCCAAAATCA 16703  
QY 149 IleLysGlyValTyrValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHis 168  
DB 16702 ATAAATGGAACGTTCTTACCGTAATACGCCAGAGCTCAAGTTACCCCAAAAATTCAC 16643  
QY 169 HisPhePheAspGlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSer 188  
DB 16642 ACCTTCATAGATGGAGATGGAATGCTCTCCACTTCGAAGATGGTCATGTCGAC 16583  
QY 189 TyrAlaCysArgPheThrGlnThrAsnArgPheValGlnGluAspGlnLeuGlyArgPro 208  
DB 16582 TTCATCAGTCGCTGGGTAAACCCGCTCGATTACGGCGGACGACTAGCGGAAATTCG 16523  
QY 209 ValPheProLysAlaIleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeu 228  
DB 16522 CTATTGGCATGTACAGAAACCCCTATACCGACGACACCAAGTGTAAAGGACTA----- 16469  
QY 229 PheTyrAlaArgAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsn 248  
DB 16468 -----GACCGCACCGCTTCCCAAT 16451  
QY 249 AlaGlyLeuValTyrPheAsnGlyArgLeuLeuAlaMetSerGluAspAspLeuProTyr 268  
DB 16450 ACAAGCATCATTAGCCATCACGCGAAGGTGTGGGGTGAAGAGACGGCTACCGTAC 16391  
QY 269 GlnValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGly 288  
DB 16390 -----GAATCGATCCTCGT---ACACTTGAACCTCGGGACACTTCGACTACACGCG 16340  
QY 289 GlnLeuGluSer---ThrMetIleAlaHisProLysValAspProGluSerGlyGluLeu 307  
DB 16339 CAAGTTACCGCAAAACCCACACCGCCCATCCAAAATATGACCCGAAACGGGTGACTTG 16280  
QY 308 PheAlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerPro 327  
DB 16279 TTG-----TTCTTCGGTTTCGGCAGCT 16259  
QY 328 AspGlyThrLysSerProAspVal----- 335  
DB 16258 AAGGGCAAGCAACTCCAGACATGGCCTATTACATTGCGACACGACGCGCAAGGTGACA 16199  
QY 336 ---GluIleGlnLeuAspGlnPro-----ThrMetMetHisAspPheAlaIleThrGlu 352  
DB 16198 CATGAAACTTGGTTTGAGCAGCCCTATGGCGCACTTCATGCGACGACTTGGCATTACCCGA 16139  
QY 353 AsnPheValValProAspGlnGlnValPheLysLeuProGluMetIleArgGly 372  
DB 16138 AATGTGTCATTTCCCAATTATGCGGCGCACCAACAGCGCTG---TCCCGCTCAAGGGG 16082  
QY 373 GlySerProVal---ValTyrAspLysAsnLysValAlaArgPheGlyIleLeuAspLys 391  
DB 16081 AAACACCAATTATATGTGGGAGCGGAACCTGGGCGAGCTACATTGGCGTACTC----- 16028

```
QY 392 TyrAlaGluAspSerAsnIleValThrPheAlaProAspCysPheCysPheHis 411
Db 16027 CGCGGGCCAGGCGAGTCTGATTCGTCGCTCAAGCAGCGGCTCTGGGTATTTTCAT 15968
QY 412 LeuTrpAsnAlaTrpGluGluProGluThrAspGluValValIleGlySerCysMet 431
Db 15967 GTTGTGAATGCTTGGGAAGTCGGAACCAAGATTATATATCGACCTTATGGAAGTGAATC 15908
QY 432 ThrPro-----ProAspSer----- 436
Db 15907 CTGCGGTTCCTCCCTCCCACTCAAAACCAACACCTTCGCCCTCGAAGAGCGGTACCA 15848
QY 437 -----IlePheAsnGluSerAspGluAsnLeuLysSerVal 448
Db 15847 CGCCTGACTCGTTGGAAATTCGCTCATAGCAGCAGCAGCAGATCAAGCGAACCCGG 15788
QY 449 LeuSerGluIleArgLeuAsnLeuLysThrGlyGluSerThr----- 462
Db 15787 CTACACGATTTCTTTGCGAAATGCCAATCGAATCTCTGTCGGCCCTGCATGCAAC 15728
QY 463 -----ArgArgProIleIleSerAsnGluAspGln 472
Db 15727 CGCTATGCTTTATGGGGTGGAGCATCCACGCAACCACTTGGCATCAGCAGCGGAG 15668
QY 473 GlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPhe 492
Db 15667 AAGATA-----TTC 15659
QY 493 AlaTyrLeuAlaLeuAlaGluProTrpProLysValSerGlyPheAlaLysValAspLeu 512
Db 15658 GCGTACAACTCACTCGC-----ATCTGG-----GACAAC 15629
QY 513 ThrThrGlyGluValLysLysHisLeuTyrGlyAspAsnArgTyrGlyGlyGluProLeu 532
Db 15628 CACCGAGGTGCTACGACCTCTGTAATCCCGCGAAGCTCGCGCGCCGCGAGCGCGCC 15569
QY 533 PheLeuPro---GlyGluGlyGlyGluGluAspGluGlyTyrIleLeuCysPheValHis 551
Db 15568 TTGCTCCCTAGAGTCGCGACCGCCCGCGAGGTGTCGCGGTACTTCTGACCGTGTGTTGTT 15509
QY 552 AspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaValSerLeuGlu----- 569
Db 15508 CGCCTCGATGAAATCGCAGCATCTGTGTAATCTCGACACTCAAGACATCCAGTCTGTT 15449
QY 570 ValGluAlaThrValLysLeuProSerArgValProTyrGlyPheHisGlyThrPheIle 589
Db 15448 CCGGTGGCAACCACTCAAGCTGCCATTCGCGCTAAGGCGCGCTCTCCATGCTGCTGGTA 15399
RESULT 13
US-08-976-063C-21/c
; Sequence 21, Application US/08976063C
; Publication No US20020182697A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Steinbuechel; Horst Priefert; Jurgen Rabenhorst
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF
; TITLE OF INVENTION: CONFERYL ALCOHOL, CONFERYLALDEHYDE, FERULIC ACID, VANILLIN A
; NUMBER OF INVENTION: ACID AND THEIR USE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESS: SPRUNG KRAMER SCHAEFER & BRISCOE
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; COMPUTER: HP VECTRA
; OPERATING SYSTEM: DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976.063C
; FILING DATE: 21-NOV-1997
```

```
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 196 49 655.1 (Germany)
FILING DATE: 29-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: Bayer 9998-CAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
TELEX:
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 1518 base pairs
TYPE: nucleic acid
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: complement (4..1518)
OTHER INFORMATION: /product=
OTHER INFORMATION: "Lignostilben-Dioxygenase"
OTHER INFORMATION: /gene="lsd"
US-08-976-063C-21
Alignment Scores:
Pred. No.: 6.68e-37 Length: 1518
Score: 411.50 Matches: 137
Percent Similarity: 40.68% Conservative: 79
Best Local Similarity: 25.80% Mismatches: 192
Query Match: 13.06% Indels: 123
Gaps: 19
US-09-758-269-6 (1-599) x US-08-976-063C-21 (1-1518)
QY 118 AspProSerValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgArg 137
Db 1497 GACCG-----CAATAGTAGGACACACTTCTCCACCGTATAGAGCAGACTTGTTC 1444
QY 138 AsnLeuProValValGlyLysLeuProAspSerIleLysGlyValTyrValArgGlnGly 157
Db 1443 GATCTAGAGTTGACGCGCAATCCCAAAATCAATAATCGAATCGTCTACCGTAATACG 1384
QY 158 AlaAsnProLeuHisGluProValThrGlyHisHisPhePheAspGlyAspGlyMetVal 177
Db 1383 CCAGAGCCTCAAGTTACCCCAAAATTCACACCTTCATAGATGAGATGGAATGGCC 1324
QY 178 HisAlaValLysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsn 197
Db 1323 TCTGCTTCCACTTCGAAGATGTCATGTCGACTTCATCATCGCTGGGTAAACCGCT 1264
QY 198 ArgPheValGlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeu 217
Db 1263 CGATTACGCGCGGACGATAGCGCGAATTCGCTATTGGCATGTACAGAAACCCCTAT 1204
QY 218 HisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaGlyIle 237
Db 1203 ACCGACGACACCACTGTAAAAGGACTA----- 1177
QY 238 ValAspProAlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArg 257
Db 1176 -----GACCGCAGCGTTGCCAATACAGCATCATTAGCCATCAGCGCAAG 1132
QY 258 LeuLeuAlaMetSerGluAspLeuProTyrGlnValGlnIleThrProAsnGlyAsp 277
Db 1131 GTGCTGGCGGTGAAGGAGACGGCTACCGTAC-----GAACCTGGATCTCCT--ACA 1081
QY 278 LeuLysThrValGlyArgPheAspGlyGlnLeuGluSer---ThrMetIleAla 296
Db 1131 GTGCTGGCGGTGAAGGAGACGGCTACCGTAC-----GAACCTGGATCTCCT--ACA 1081
```

Db 1080 CTTGAACTCGGACACACTTCGACTACGACGGCCAGTTTACACGCAAAACCCACACGCC 1021  
Qy HisProLysValAspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSer 316  
Db CATCCAAATATGACCCGCAACCGGTCGACTGTGTTG----- 985  
Qy LysProTyrLeuLysTyrPheArgPheSerProAspGlyThrLysSerProAspVal--- 333  
Db -----TTCTTCGGTTCGACACTAAGGGCGAAGCAACTCCAGACATGGCC 940  
Qy 336 -----GluLeuGlnLeuAspGlnPro--- 342  
Db TATTACATTGTCGACACGACGCAAGGTGACATGAACAACTTGTTTGACGACCCCTAT 880  
Qy 343 ---ThrMetMetHisAspPheAlaTleThrGluAsnPheValValProAspGlnGln 361  
Db GCGCATTCATGACGACTTTCGCAATACCGAAATGTCCTCCATTTCCCAATTATGCGG 820  
Qy 362 ValValPheLysLeuProGluMetIleArgGlyGlySerProVal---ValTyrAspLys 380  
Db GCCACCAACAGCGCTG---TCCCGCTCAAGCGCAACAGCAATTTATATGTTGGAGCGG 763  
Qy 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLys 400  
Db GAATCGGCGACGACTACATTGGCGTATC-----GCGCCGCGCCAGGGCAGTCTGATTGCG 709  
Qy 401 TrpIleAspAlaProAspCysPheCysPheHisLeuTrpAsnAlaTrpGluGluProGlu 420  
Db TGGCTCAAGGACCGCGCTCGGTATTTTCATGTTGTGAATGCTTGGGAAGTCGGAACC 649  
Qy 421 ThrAspGluValValIleGlySerCysMetThrPro-----ProAspSer--- 436  
Db AAGATTATATGACCTTATGAAAGTGAATCTCGCGTTCCTCCCACTCAAA 589  
Qy 437 -----Ile 437  
Db 588 AACCAACCTTCGCCCCGAGAAACCGCTACCAACCGCTGACTCGTTGGGAATTTGACCTC 529  
Qy 438 PheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLys 457  
Db 528 GATGACGAGCGGACGATCAAGCGAACCAGGCTACACGATTTCTTTCGGAATAATGCCA 469  
Qy 458 ThrGlyGluSerThr----- 462  
Db 468 ATCATGGAATTTCTGTTGCCCTCGAATGCAACCGCTATGGCTTTATGGGGTGGACGAT 409  
Qy 463 ---ArgArgProIleIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMetVal 481  
Db 408 CCACGCAACCACTTCGCGCATCAGCAGCGCGGAGAGATA----- 370  
Qy 482 AsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluProTrp 501  
Db 369 -----TTCGGCTACAACTCACTCTCGGC---ATCTGG 343  
Qy 502 ProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysHisLeu 521  
Db 342 -----GACAAACCCAGCGAGTACGACCTCTGTGATC 310  
Qy 522 TyrGlyAspAsnArgTyrGlyGluProLeuPheLeuPro---GlyGluGlyGlyGlu 540  
Db 309 TCCGCGAAGCCTCGCGCGCCAGAGCGCGCTTCGTCCTCTGAGAGTCCGACCGCCGCC 250  
Qy 541 GluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLysSerGluLeu 560  
Db 249 GAAGGTGACGGTACTTGTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 190  
Qy 561 GlnIleValAsnAlaValSerLeuGlu-----ValGluAlaThrValLysLeuProSer 578  
Db 189 GTAATTCGACACTCAAGACATCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 130  
Qy 579 ArgValProTyrGlyPheHisGlyThrPheIle 589  
Db 129 CGGCTAAGGCGCGCTCTCCATGGCTGCTGGTGA 97

## RESULT 14

US-09-878-574-2872  
; Sequence 2872, Application US/09878574  
; Patent No. US20020110548A1  
; GENERAL INFORMATION:  
; APPLICANT: Byrum, Joseph R.  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Thompson, Michael D.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(15401)B  
; CURRENT APPLICATION NUMBER: US/09/878,574  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/333,535  
; PRIOR FILING DATE: 1999-06-14  
; NUMBER OF SEQ ID NOS: 15775  
; SEQ ID NO 2872  
; LENGTH: 320  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)-(320)  
; OTHER INFORMATION: unsure at all n locations  
; OTHER INFORMATION: Clone ID: LIB3028-019-Q1-B1-B11  
; US-09-878-574-2872

Alignment Scores:  
Pred. No.: 4,596-33 Length: 320  
Score: 369.00 Matches: 68  
Percent Similarity: 80.73% Conservative: 20  
Best Local Similarity: 62.39% Mismatches: 15  
Query Match: 11.71% Indels: 6  
DB: 10 Gaps: 2

US-09-758-269-6 (1-599) x US-09-878-574-2872 (1-320)

Qy 447 SerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGluSerThrArgArgProIle 466  
Db 2 AGCGTTTTTAAAGAAAGTAAAGCTGACATGAGAGAGGAGGAGGAGGAGGAGGAGGAGG 51  
Qy 467 IleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeu 486  
Db 62 GTG-----GAGGAATGAACCTGGAGGCGAGTGGTGAACAGGAAAGGTTG 109  
Qy 487 GlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluProTrpProLysValSerGly 506  
Db 110 GGGAGAAAAACACGGTTCGCATATTTTGCATAGCGGAACCGTGGCCGAGGTGCGGG 169  
Qy 507 PheAlaLysValAspLeuThrThrGlyGluValLysHisLeuTyrGlyAspAsnArg 526  
Db 170 GTGCGAAGGTGGACCTGGAGAGTGGGAGAGTGAAGAGGCACGAGTACGAGAGAGAAG 229  
Qy 527 TyrGlyGlyGluProLeuPheLeuPro-----GlyGluGlyGlyGluGluGly 544  
Db 230 TTGCGTGGAGAGCAATCTTCTTCCCAACACGCTGGAGGAATGGGAATGAGATGAANG 289  
Qy 545 TyrIleLeuCysPheValHisaspGlu 553  
Db 290 TACGTATGCTTTTGTGATGACGAG 316

## RESULT 15

US-09-770-696-172/c  
; Sequence 172, Application US/09770696  
; Patent No. US2001004940A1  
; GENERAL INFORMATION:  
; APPLICANT: Gorlach, Jörn  
; APPLICANT: An, Yong-Qiang  
; APPLICANT: Hamilton, Carol M.  
; APPLICANT: Price, Jennifer L.  
; APPLICANT: Raines, Tracy M.  
; APPLICANT: Yu, Yang

APPLICANT: Rameaka, Joshua G.  
APPLICANT: Page, Amy  
APPLICANT: Matthew, Abraham V.  
APPLICANT: Ledford, Brooke L.  
APPLICANT: Woessner, Jeffrey P.  
APPLICANT: Haas, William David  
APPLICANT: Garcia, Carlos A.  
APPLICANT: Krickler, Maja  
APPLICANT: Slader, Ted  
APPLICANT: Davis, Keith R.  
APPLICANT: Allen, Keith  
APPLICANT: Hoffman, Neil  
APPLICANT: Hurban, Patrick  
TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
FILE REFERENCE: thaliana  
CURRENT APPLICATION NUMBER: US/09/770,696  
CURRENT FILING DATE: 2001-01-26  
PRIOR APPLICATION NUMBER: 60/178,278  
PRIOR FILING DATE: 2000-01-27  
NUMBER OF SEQ ID NOS: 911  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 172  
LENGTH: 200  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (1)..(200)  
OTHER INFORMATION: n = A,T,C or G  
US-09-770-696-172

Alignment Scores:  
Pred. No.: 9.82e-31 Length: 200  
Score: 346.00 Matches: 65  
Percent Similarity: 98.48% Conservative: 0  
Best Local Similarity: 98.48% Mismatches: 1  
Query Match: 10.98% Indels: 0  
DB: 9 Gaps: 0

US-09-758-269-6 (1-599) x US-09-770-696-172 (1-200)

Qy	265	AspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPhe	284
Db	199	GATTACCTTACCAGTTCAGATCCTCCCAATGGAGATTATAAACCGTTGGTCGGTTC	140
Qy	285	AspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysValAspProGluSer	304
Db	139	GATTTTGATGACAAATTAGAAATCCCAATGATTGCCCCGCCGAAAGTCGACCCGGAATCC	80
Qy	305	GlyCluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArg	324
Db	79	GGTGAAACNNITCGGTTTAAAGTCAGCGTCGTTTCAAAGCCTTACCTAAATAACTTCCGA	20
Qy	325	PheSerProAspGlyThr	330
Db	19	TTCTCACCAGCGGAACT	2

Search completed: November 17, 2003, 00:51:50  
Job time : 574 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 16, 2003, 20:21:32 ; Search time 96 Seconds

(Without alignments)  
2754.047 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150

Sequence: 1 MASTATAAVSGRWLGNGHNT.....VPYGFHCTFFIGADDLAKQVV 599

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh

-O=/cn2\_1/usptc.spool/US09758269/runat.14112003.192309.25781/app.query.fasta.1.775  
-DB=Issued\_Patents\_NA -OPMT=fastap -SUFFIX=rn1 -XINMATCH=0.1 -LOEFCI=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09758269 @CN 1.1 56 @runat.14112003.192309.25781 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued\_Patents\_NA:\*  
1: /cn2\_6/prodata/1/ina/5A COMB.seq:\*  
2: /cn2\_6/prodata/1/ina/5B COMB.seq:\*  
3: /cn2\_6/prodata/1/ina/6A COMB.seq:\*  
4: /cn2\_6/prodata/1/ina/6B COMB.seq:\*  
5: /cn2\_6/prodata/1/ina/PCUS COMB.seq:\*  
6: /cn2\_6/prodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	494.5	15.7	4403765	3	US-09-103-840A-2
2	494.5	15.7	4411529	3	US-09-103-840A-1
3	475.5	15.1	4403765	3	US-09-103-840A-2
4	475.5	15.1	4411529	3	US-09-103-840A-1
5	419	13.3	32679	4	US-08-976-063E-1
6	411.5	13.1	1518	4	US-08-976-063E-21
7	211	6.7	2629	1	US-08-200-807-1
8	211	6.7	2629	1	US-08-488-305A-1
9	210	6.7	1724	3	US-09-385-259-1
10	210	6.7	1724	4	US-09-645-370-1
11	186	5.9	36063	4	US-08-311-731A-140
12	127.5	4.0	4242	4	US-09-252-991A-7056

13	127.5	4.0	10023	4	US-09-252-991A-6997	Sequence 6997, Ap
14	123.5	3.9	5163	3	US-08-700-651-1	Sequence 1, Appl
15	123.5	3.9	5163	3	US-08-928-361B-4	Sequence 4, Appl
16	123.5	3.9	5163	4	US-09-588-995A-4	Sequence 4, Appl
17	123.5	3.9	5318	3	US-08-700-651-2	Sequence 2, Appl
18	123.5	3.9	5318	3	US-08-928-361B-3	Sequence 3, Appl
19	123.5	3.9	5318	4	US-09-588-995A-3	Sequence 3, Appl
20	116.5	3.7	5511	3	US-08-928-361B-2	Sequence 2, Appl
21	116.5	3.7	5511	4	US-09-588-995A-2	Sequence 2, Appl
22	116.5	3.7	7334	3	US-08-928-361B-1	Sequence 1, Appl
23	116.5	3.7	7334	4	US-09-588-995A-1	Sequence 1, Appl
24	115	3.7	1389	4	US-09-252-991A-11721	Sequence 11721, A
25	115	3.7	2322	4	US-09-252-991A-11519	Sequence 11519, A
26	115	3.7	47981	4	US-09-679-279-1	Sequence 1, Appl
27	110	3.5	1835	4	US-09-252-991A-2493	Sequence 2493, Ap
28	106	3.4	1839	3	US-09-461-697-76	Sequence 76, Appl
29	106	3.4	1825	3	US-09-461-697-75	Sequence 75, Appl
30	106	3.4	1953	4	US-09-252-991A-759	Sequence 759, App
31	105.5	3.3	1416	4	US-09-107-532A-2051	Sequence 2051, Ap
32	105.5	3.3	4508	5	PCT-US93-06251-34	Sequence 34, Appl
33	105	3.3	1172	3	US-08-861-774E-17	Sequence 17, Appl
34	104.5	3.3	1389	4	US-09-328-352-3002	Sequence 3002, Ap
35	104	3.3	1341	4	US-09-350-756-2	Sequence 2, Appl
36	103	3.3	1479	4	US-09-252-991A-4180	Sequence 4180, Ap
37	103	3.3	1575	4	US-09-252-991A-2009	Sequence 2009, Ap
38	103	3.3	1581	4	US-09-252-991A-654	Sequence 654, App
39	103	3.3	3102	4	US-09-252-991A-4429	Sequence 4429, Ap
40	102.5	3.3	2607	2	US-08-907-166-1	Sequence 1, Appl
41	102.5	3.3	2607	4	US-09-391-340-1	Sequence 1, Appl
42	102.5	3.3	3257	4	US-09-585-173B-39	Sequence 39, Appl
43	102.5	3.3	14672	4	US-08-961-527-111	Sequence 111, App
44	102	3.2	875	4	US-09-252-991A-12072	Sequence 12072, A
45	102	3.2	3396	4	US-09-252-991A-14676	Sequence 14676, A

#### ALIGNMENTS

#### RESULT 1

US-09-103-840A-2  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

#### Alignment Scores:

Pred. No.: 4.22e-39  
Score: 494.50  
Percent Similarity: 40.18%  
Best Local Similarity: 26.39%  
Query Match: 15.70%  
DB: 3  
Length: 4403765  
Matches: 180  
Conservative: 94  
Mismatches: 251  
Indels: 157  
Gaps: 27

US-09-758-269-6 (1-599) x US-09-103-840A-2 (1-4403765)



```

; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.: 4,23e-39 Length: 4411529
Score: 494.50 Matches: 180
Percent Similarity: 40.18% Conservative: 94
Best Local Similarity: 26.39% Mismatches: 251
Query Match: 15.70% Indels: 157
Db: 3 Gaps: 27

US-09-758-269-6 (1-599) x US-09-103-840A-1 (1-4411529)
Qy 1 MetAlaSerPheThrAlaThr---AlaAlaValSerGly-----Arg 13
Db 749677 ATGCCACAGGTCGGCGACCGGATCGCGGTCGTGCGGACCGTCGACGCGACATCGAAC 749736
Qy 14 TrpLeuGlyGlyAsnHisThrGlnProProLeuSerSerSerGlnSerSerAspLeuSer 33
Db 749737 TGGCGTAGCCCTCTCTCGGTATGCGCGCGATGATCGCGCATCTCGCGCATCCCGCGAAATGGGTG 749796
Qy 34 TyrCysSer-----SerLeuProMetAlaSerArgValThrArgLysLeuAsnVal--- 50
Db 749797 TACACCGCATCGTCGAGGTGCTCTCGCGCGCCACCTTCGCGGTCTGCGAGCGCTCG 749856
Qy 51 -----SerSerAlaLeuHisThrProProAlaLeuHisPheProLysGlnSerSer 67
Db 749857 GCGCGTGTATCGTCGACGATCGACGCGCGCGTGC-----AGCAGTCTCGTCG 749904
Qy 68 AsnSerProAlaIleValVal-----LysProLysAlaLysGluSerAsn 82
Db 749905 CGAACACCGGTCTGCGAGGTATCTCTCCATGTTCTCACCAAGGGGTACCGTTCCAAT 749964
Qy 83 ThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAlaAlaGluGly 102
Db 749965 ATCAGTGAATAACAATGTTATAGAGATCGCATGACCCGACACGACGCGCGAA--- 750021
Qy 103 PheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSerValGln 122
Db 750022 -----TCCAAACCCATATCTCGAG 750042
Qy 123 IleAlaGlyAsnPheAlaProValAsnGluGlnProValArgArgAsnLeuProValVal 142
Db 750043 -----GGCTTCCTCGCGCGGTGAGCACCGAGGTAATGCGACCGCATCTCGCGTCAAC 750096
Qy 143 GlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnProLeuHis 162
Db 750097 GCGCGCATTCGGAACACCTCGACGCGGTATCTGCGTAAACGCGCCCAACCGCGTCCG 750156
Qy 163 Glu-----ProValThrGlyHisHisPhePheAspGlyAspGlyMetValHisAlaVal 180
Db 750157 GAGGTGCGACCGGCCACCC---TACCACCTGGTTTACCGCGCGAGCCCATGTCGACGGAGTC 750213
Qy 181 LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal 200
Db 750214 GCGTTCGCGACGCGGAAGGCC-----CGCTGGTAT 750243
Qy 201 GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGlu----- 216

```



```
QY 519 LysHisLeuTyrGlyAspAsnArgTyrGlyGluProLeuPheLeuProGlyGluGly 538
Db 751276 GTCCGCTCGCTGATCCGACCTGCTGATCGCGAGATGGTGTTCGTCGCGAACCCGTCG 751335
QY 539 GlyGlu---GluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLys 557
Db 751336 GCGCGTCGAGAGATGACGGGATTCCTATGGCTACGGCTGGCACCAGCGCGGCGGACGAA 751395
QY 558 SerGluLeuGlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuPro 577
Db 751396 GCCAGCTGCTCTTCTGTCGATCCAGACTCTCGAGTCGATCGCCACCGCTGCACCTGCCA 751455
QY 578 SerArgValProTyrGlyPheHisGlyThrPhe-----IleGly 590
Db 751456 CAGCGGTGCCGATGGCTTCCAGCGCAACTGGGCGCGACCACTGACGCGGCTCGGG 751515
QY 591 AlaAsp 592
Db 751516 TGGCAT 751521

RESULT 3
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 5,52e-37 Length: 4403765
Score: 165 Matches: 165
Percent Similarity: 40.35% Conservative: 90
Best Local Similarity: 26.11% Mismatches: 255
Query Match: 15.10% Indels: 123
DB: 3 Gaps: 23

US-09-758-269-6 (1-599) x US-09-103-840A-2 (1-4403765)
QY 17 GlyAsnHisThrGlnProProLeuSerSerSerClnSerSerAspLeuSerTyrCysSer 36
Db 1018941 GCGCGTACCATCAACCCAGCGGAGGATGATCGGTGGCGGACATCCGG-----TCG 1018888
QY 37 SerLeuProMetAlaSerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThr 56
Db 1018887 GGGCTTCGGGTGTCGGATGCTGCTCGACGCGGGCCAAACAGGTGAGCGGATTCGCGGTG 1018828
QY 57 ProProAlaLeuHisPheProLysGlnSerSerSerSerProAlaIleValLysPro 76
Db 1018827 ACTATCAGGTGCGAGACCGCCGACGCTGCGCAGCTGAATTCGGTGGCAGCAGGCCCA 1018768
QY 77 LysAlaLysGluSerAsnThrLysGlnMetAsnLeuPheGlnArgAlaAlaAla 96
Db 1018767 CCACGGTCAGTTTCG-----TTGTCAGCACACAC 1018738
```

Db 1017792 -----TTCTACGAAGCGATCCGACGCCACTTGACACCGGAGAACGAA 1017749  
Qy 443 -----GluAsnLeuLysSerValLeuSerGlu1 452  
Db 1017746 GTGGAGAAGCTGTTCCGTTCTCGGCTCTGGATCGCTCGACTCGGCTCATCGTG 1017689  
Qy 452 eArgLeuAsnLeuLysThrGlyGluSerThrArgArgProIleIleSerAsnGluAspG1 472  
Db 1017688 GCGGCTCAATATGTTGACCGGG-----GCAGTCCACGAGGA 1017653  
Qy 472 nGlnValAsn-----LeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLy 489  
Db 1017652 GCAACTGTCCAGATCCATCACCGAGTTCGGAACCATCAACCGCGATTACCGGCCACGAG 1017593  
Qy 489 sThrLysPheAlaTyrLeuAlaLeuAlaGluPro---TTPProLysValSerGlyPheAl 508  
Db 1017592 CTACCGTTACCTATGCCGTACCGGCAACCGAGTGTGTTCTGTTCCGACGACTGGT 1017533  
Qy 508 aLysValAspLeuThrThrGlyGluValLysLysHisLeuTyrGlyAspAsnArgTyrG1 528  
Db 1017532 TAAGCACGATCTGCTACCGGCAACCGAGTGTACTCTGTTCTGTTGACGCGCTCTACGG 1017473  
Qy 528 yGlyGluProLeuPheLeuProGlyGluGlyGlu---GluAspGluGlyTyrIleLe 547  
Db 1017472 AAGTGAGACCGGATGGCTCCACGGTGGGAGCAGCGCCGCGAGGACGACGGCTATCTGGT 1017413  
Qy 547 uCysPheValHisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaValse 567  
Db 1017412 CACCCTCACCGACATGAACGACGACGATCGTATTGCTGTTGCTGTTTCGACGCGGCGG 1017353  
Qy 567 rLeuGluValGluAlaThrValLys-----LeuProSerArgValProTyrGlyPheHi 585  
Db 1017352 CCCCAGGATGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1017293  
Qy 585 sGlyThrPheLeuGlyAlaAspLeuAlaLys 596  
Db 1017292 TTCGCGGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1017259  
RESULT 4  
US-09-103-840A-1/c  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R. M.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv  
US-09-103-840A-1  
Alignment Scores:  
Pred. No.: 5,53e-37 Length: 4411529  
Score: 475.50 Matches: 165  
Percent Similarity: 40.35% Conservative: 90  
Best Local Similarity: 26.11% Mismatches: 255  
Query Match: 15.10% Indels: 123  
DB: 3 Gaps: 23  
US-09-758-269-6 (1-599) x US-09-103-840A-1 (1-4411529)  
Qy 17 GlyAsnHisThrGlnProLeuSerSerGlnSerSerAspLeuSerTyrCysSer 36

Db 1018939 GCGGCTACCCATCAACCCAGCGGAGGACTGATCGTGGCGGACATCCGG-----TCG 1018886  
Qy 37 SerLeuProMetAlaSerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThr 56  
Db 1018885 GGGCTTCGGGTGCGGATGCTGCTCGACGCGGCAACAGCTGAGCGGCTATTCGGCGTG 1018826  
Qy 57 ProProAlaLeuHisPheProLysSerSerAsnSerProAlaIleValValLysPro 76  
Db 1018825 ACTATCAGTCCAGAACCGGAGCGTTCGGACGCTGAATTCGTTGGTGGCAGCAGGCA 1018766  
Qy 77 LysAlaLysGluSerAsnThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAla 96  
Db 1018765 CCACGTCAGTTTCG-----TTGTCAGCACCACAC 1018736  
Qy 97 LeuAspAlaAlaGluGlyPheLeuVal-SerHisGluLysLeuHisProLeuProLys-- 115  
Db 1018735 GAGGATCGTGAATAGACATCAGCATTCGCGCAAGTACTTGTGACACCTTCCCGAAGAC 1018676  
Qy 116 -ThrAlaAspProSerValGlnIleAlaGlyAsnPheAlaProValAsnGluInProVa 135  
Db 1018675 GACGACACCCCTATCGCACCGGCTCGTGGGACACACAGACCCAGCATGGATGCCAC 1018616  
Qy 135 lArgArgAsnLeuProValValGlyLeuProAspSerIleLysGlyValTyrValar 155  
Db 1018615 GACCTG-ACC---ACCGTGACAGCGGAAGTCCCGCCGACCTGGACGCGCATCTACTCGG 1018560  
Qy 155 gAsnGlyAlaAsnProLeuHisGluProValThrGlyHisPheHisPheAspGlyAspG1 175  
Db 1018559 CAACACCGAGAACCCGCTACACCCGCGATTCGCGACCTACCAACCTTCGATGGCGACGG 1018500  
Qy 175 yMetValHisAlaValLysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrG1 195  
Db 1018499 CATGATCCATCGTTCGGCTTCGGTGATGGAAAGCCCTTACCGCAACCGATTATTTCG 1018440  
Qy 195 nThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPheProLysAlaIleG1 215  
Db 1018439 CACCGATGATTCCTTGGCCGAGAACGAGCGCGCGCGCGCTGTGGCGC----- 1018391  
Qy 215 yGluLeuHisGlyHisThrGlyLeuAlaArgLeuMetLeuPheTyrAlaArg----- 232  
Db 1018390 -----GGTCTGGCAACCGGTGCACTGGCGCAAGCGGGAACACGG 1018350  
Qy 233 ----AlaAlaAlaGlyIleValAspProAlaHisGlyThrGlyVal---AlaAsnAlaG1 250  
Db 1018349 CTGGGCGCTCGTGGCTCATGAAGACGCGTCGACGACCGACGCTCATCTCCCGCAGG 1018290  
Qy 250 yLeuVal-----TyrPheAsnGlyArgLeuLeuAlaMetSerGluAspAspLe 266  
Db 1018289 TATCGCGCTGACTAGTCTTCTACAGTGGCGCGCATCTGTATCGATCGAC----- 1018241  
Qy 266 uProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPh 286  
Db 1018240 -CCGTACTCGGCCAATACGCTC-----GGCAAGGAGAGCTGGCACGGAAGGTTTCCGTT 1018188  
Qy 286 eAspGlyGlnLeuGluSerThrMetIleAlaHisProLysValAspProGluSerGlyG1 306  
Db 1018187 CGACTGGGCGGTGCTG-----GCACATCCGAGAGGTAGACAAAGACCGCGCA 1018140  
Qy 306 uLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSe 326  
Db 1018139 ACTGTTCTTCTCACTACAGC---AAGCAAGAGCGCTATATCGCTACGCGCTTCGCA 1018083  
Qy 326 rProAspGlyThrLysSerProAspValGluIleGlnLeuAspGlnProThrMetMetHi 346  
Db 1018082 CCAGAACAATAGACTTGTGCGATCATCTGATTCGCTGCGCGCGCGCGCTACCGCA 1018023  
Qy 346 sAspPheAlaIleThrGluAsnPheValValProAspGlnGlnValValPheLysLe 366  
Db 1018022 TGACATGGCGTTTACCGGAAATTTACGTAATCTCTCAAGAT----- 1017983  
Qy 366 uProGluMetIleArgGlyGlySerProValValTyrAspLysAsnLysVal----- 383

Db 1017982 -----TTTCCACTGTTCTGGGATCCAGGCTGCTCGAGCGCA 1017945  
QY 384 -----AlaArgPheGlyIleLeuAspLysTy 392  
Db 1017944 COTGACCTACACGCTTCTATCCGAGATTCATCTCGGTTCGGGTGTGT----- 1017893  
QY 392 rAlaGluAspSerSerAsnIleLysTrpIleAspAlaProAspCysPheCysPheHisLe 412  
Db 1017892 -GCTCGCGAGGCAACGACATCTGCTGTTCGAAGCCGATCAACGTTCTGTGTGCACTT 1017834  
QY 412 uTrpAsnAlaTrpGluGluProGluThrAspGluValValIleGlySerCysMetTh 432  
Db 1017833 CACCAACGCTACGAG-----CAGGGGACGAGATCGTGTCTCGACGCG----- 1017791  
QY 432 rProProAspSerIlePheAsnGluSerAsp----- 442  
Db 1017790 -----TTCTACGAAGGCGATCCGAGCGACCTTGACACCGGAGAACGAA 1017747  
QY 443 -----GluAsnLeuLysSerValLeuSerGluI 452  
Db 1017746 GTGGGAGAGCTGTTTCGGTTCCTCGCTCGGTGTCAGTCCCGCTACATCGGTG 1017687  
QY 452 eArgLeuAsnLeuLysThrGlyGluSerThrArgArgProIleIleSerAsnGluAspG 472  
Db 1017686 GCGGCTCATATGGTCACCGG-----GCAGTCCACGAGGA 1017651  
QY 472 nGlnValAsn-----LeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLy 489  
Db 1017650 GCAACTGTCGAGTCCATCCAGGATTCGGAACCATCAACGCGGATTACGCGCCAGAG 1017591  
QY 489 sThrLysPheAlaTyLeuAlaLeuAlaGluPro---TrpProLysValSerGlyPheAl 508  
Db 1017590 CTACCGTTACACTATGCGGTACCGGCAACCGAGTTGTTCTGTTCCGACGAGTGGT 1017531  
QY 508 aLysValAspLeuThrThrGlyGluValLysLysHisLeuTyGlyAspAsnArgTyG 528  
Db 1017530 TAAGCAGCATCTCTCACCGGCAACCAACGAGTGTACTCTGTTGCGGTGACGGGTCTACGG 1017471  
QY 528 yGlyGluProLeuPheLeuProGlyGlyGlyGlu---GluAspGluGlyTyIleLe 547  
Db 1017470 AAGTAGACCGCGATGCTCCAGGTGGGCACGACGCGCGGAGGACGCGGTATCTGT 1017411  
QY 547 uCysPheValHisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaValSe 567  
Db 1017410 CACCTCACCACGACATGAACGACGACGATCTATTCGCTGTTTTCGACGCGCGCG 1017351  
QY 567 rLeuGluValGluAlaThrValLys-----LeuProSerArgValProTyGlyPheHi 585  
Db 1017350 CCGCGCGATGGCCCGATATGCAAGCTTGCACTGCGGACGACGATTTTCCAGCGGCGCA 1017291  
QY 585 sGlyThrPheIleGlyAlaAspAspLeuAlaLys 596  
Db 1017290 TTCGCGGTGGTGGCGGCGCGGAGTTGCGTCGC 1017257

RESULT 5  
US-08-976-063E-1/c  
; Sequence 1, Application US/08976063E  
; Patent No. 6524831  
; GENERAL INFORMATION:  
; APPLICANT: Steinbuechel, Alexander  
; APPLICANT: Priefert, Horst  
; APPLICANT: Rabenhorst, Jürgen  
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL  
; TITLE OF INVENTION: ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN AND  
; TITLE OF INVENTION: VANILLIC ACID AND THEIR USE  
; FILE REFERENCE: Bayer-9998-CAO  
; CURRENT APPLICATION NUMBER: US/08/976,063E  
; CURRENT FILING DATE: 1997-11-21  
; PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY  
; PRIOR FILING DATE: 1996-11-29  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 1

/ LENGTH: 32679  
; TYPE: DNA  
; ORGANISM: Pseudomonas sp.  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (3146)..(3937)  
; OTHER INFORMATION: gene = "ORF1"  
US-08-976-063E-1  
Alignment Scores:  
Pred. No.: 2,58E-34 Length: 32679  
Score: 419.00 Matches: 160  
Percent Similarity: 37.97% Conservative: 83  
Best Local Similarity: 25.00% Mismatches: 229  
Query Match: 13.30% Indels: 168  
DB: 4 Gaps: 24  
US-09-758-269-6 (1-599) x US-08-976-063E-1 (1-32679)  
QY 19 HisThrGlnProProLeuSerSerSerGlnSerSerAspLeuSerTyrcysSerLeu 38  
Db 17011 CATACGCCGCCAGATAGTTCATTCTTGAATTTCTAAACAATAGCCAGCTCGGC 16952  
QY 39 ProMetAlaSerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProPro 58  
Db 16951 CCATTTTCA-----ATACCACCA 16934  
QY 59 AlaLeuHisPheProLysGlnSerSerAsn-----SerProAlaIleValVal 74  
Db 16933 CCATTGCTCTCACAGACATCTTCCTCACCAGCACACATGAATCG-----GTC 16883  
QY 75 LysProLysAlaLysGluSerAsnThrLys-----GlnMetAsnLeu 88  
Db 16882 AGTCCTAAACATAAATCCCAACCCGCGCAGCTTAGGCCATTAGGCCATTAGAACACAAA 16823  
QY 89 PheGlnArgAlaAlaAlaAlaLeuAspAlaAlaGluGlyPheLeuValSerHisGlu 108  
Db 16822 GGAGACGCTGCCATGGCGAGATTC----- 16799  
QY 109 LysLeuHisProLeuProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAla 128  
Db 16798 -----AACCGCACACGACCG-----CAATTAGTAGGAACACTTCTC 16763  
QY 129 ProValAsnGluGlnProValArgAsnLeuProValValGlyLysLeuProAspSer 148  
Db 16762 CCCACCGGTATAGAGGAGACTTGTTCATCTAGAGTTACGCGCGAAATCCCAAAATCA 16703  
QY 149 IleLysGlyValTyValValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHis 168  
Db 16702 ATAAATGGAACTTCTACCGTAAATACGCGAGAGCTCAAGTTACCCACACAAAATTCAC 16643  
QY 169 HisPheAspGlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSer 188  
Db 16642 ACCTTCATAGATGAGATGGAATGGCTCTCCCTTCATTCGAAGATGGTCATGTCGAC 16583  
QY 189 TyrAlaCysArgPheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgPro 208  
Db 16582 TTCATCAGTCTCGGTATTAACCGCTCGATTACGCGCGAAACGACTAGCGCGAAATCG 16523  
QY 209 ValPheProLysAlaIleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeu 228  
Db 16522 CTATTGGCATGTACAGAAACCCCTATACCGACACACCACTGTAAAGAGCTA----- 16469  
QY 229 PheTyAlaAlaAlaAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsn 248  
Db 16468 -----GACCGCACCGTTGCCAAT 16451  
QY 249 AlaGlyLeuValTyPheAsnGlyArgLeuLeuAlaMetSerGluAspAspLeuProTy 268  
Db 16450 ACAAGCATCATTAGCCATCAGCGCAAGGTGCTGCGGTGAAGGAGAGCGGCTACCGTAC 16391  
QY 269 GlnValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspGly 288

Db 16390 -----GACTGGATCTCTCGT---ACACTTGAAACTCGCGGACACACTTCGACTAGCAGCGC 16340

Qy 289 GlnLeuGluser---ThrMetIleAlaHisProLysValAspProGluSerGlyGluLeu 307

Db 16339 CAAGTTACAGCAACACACACCGCCCATCCAAATATATGACCGGAAACGGGTGACTTG 16280

Qy 308 PheAlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerPro 327

Db 16279 TTG-----TTCITTCGGTTCCGCGAGCT 16259

Qy 328 AspGlyThrLysSerProAspVal-----TTCITTCGGTTCCGCGAGCT 16259

Db 16258 AAGGCGAAGCAACTCCAGACATGCTTATACATTGTCGACAGCAGCGCAAGGTGACA 16199

Qy 336 ---GluIleGlnLeuAspGlnPro-----ThrMetMetHisAspPheAlaIleThrGlu 352

Db 16198 CATGAACTTGGTTGACAGCCCTATGCGCATTCATGACGACTTTCGCATTACCCGA 16139

Qy 353 AsnPheValValProAspGlnValValPheLysLeuProGluMetIleArgGly 372

Db 16139 AATGGTCCATTTCCTCCCAATTATCGCGCCACCAACAGCCTG---TCCCGCCTCAAGCGC 16082

Qy 373 GlySerProVal---ValTyrAspLysAsnLysValAlaArgPheGlyIleLeuAspLys 391

Db 16081 AAACAGCCCAATTATATGTGGGACCGGAACTGGCGAGCTACATTGCGGTACTC----- 16028

Qy 392 TyrAlaGluAspSerAsnIleLysTrpIleAspAlaProAspCysPheCysPheHis 411

Db 16027 CGCGCGCCAGGCGAGCTCTGATTCGTGCTCAGGCACCGCGCTCTGGATTTCAT 15968

Qy 412 LeuTrpAsnAlaTrpGluGluProGluThrAspGluValValIleGlySerCysMet 431

Db 15967 GTTGTGAATGCTTGGGAAGTCCGGAACCAAGATTATATCGACCTTATGGAAGTGAATC 15908

Qy 432 ThrPro-----ProAspSer----- 436

Db 15907 CTGCGGTCCCTTCCCACTCACAACCAACCTTCGCCCTTGAGAAAGCGGTACCA 15848

Qy 437 -----IlePheAsnGluSerAspGluAsnLeuLysSerVal 448

Db 15847 CGCCTGACTCTGGGAAATGACCTGATAGCAGCAGCAGAGATCAAGCAACCCGG 15788

Qy 449 LeuSerGluIleArgLeuAsnLeuLysThrGlyGluSerThr----- 462

Db 15787 CTACAGATTCTTTCGGAAATGCCAATCATGATGATTCTTCGTTCGCCCTGCAATGCAAC 15728

Qy 463 -----ArgArgProIleLeuSerAsnGluAspGln 472

Db 15727 CGCTATGCTTTATGGGGTGGACGATCCACGCAACCACTTCGCGCATCAGCAGCGGAG 15668

Qy 473 GlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPhe 492

Db 15667 AAGATA-----TTC 15659

Qy 493 AlaTyrLeuAlaLeuAlaGluProTrpProLysValSerGlyPheAlaLysValAspLeu 512

Db 15658 CGGTACAACTCACTCGGC---ATCTGG-----GACAAC 15629

Qy 513 ThrThrGlyGluValLysLysHisLeuTyrGlyAspAsnArgTyrGlyGlyGluProLeu 532

Db 15628 CACGAGGTGACTACGACCTCTGTTACTCCGCGGAAAGCTTCGCGCGCCAGGAGCGCGCC 15569

Qy 533 PheLeuPro---GlyGluGlyGlyGluAspGlyTyrIleLeuCysPheValHis 551

Db 15568 TCGTCCCTAGAAGTCCGACCGCCGCGGAGAGTGCAGCGGTACTGCTGACCGGTGGT 15509

Qy 552 AspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaValSerLeuGlu----- 569

Db 15508 CGCCTCGATGAAATCGCAGCGATCTGGTAAATTCGACACTCAAGACATCCAGCTCTGGT 15449

Qy 570 ValGluAlaThrValLysLeuSerArgValProTyrGlyPheHisGlyThrPheIle 589

Db 15448 CCCGTGGCAACCATCAAGCTCCCATTCGCGCTAAGGCGCGCTCTCCATGGCTCTGGGTA 15389

RESULT 6

US-08-976-063E-21/c  
; Sequence 21, Application US/08976063E  
; Patent No. 6524831  
; GENERAL INFORMATION:

; APPLICANT: Steinbuechel, Alexander  
; APPLICANT: Priefest, Horst  
; APPLICANT: Rabenhorst, Jürgen  
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL  
; TITLE OF INVENTION: ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN AND  
; TITLE OF INVENTION: VANILLIC ACID AND THEIR USE  
; FILE REFERENCE: Bayer-9998-CAO  
; CURRENT APPLICATION NUMBER: US/08/976,063E  
; CURRENT FILING DATE: 1997-11-21  
; PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY  
; PRIOR FILING DATE: 1996-11-29  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 21

; LENGTH: 1518

; TYPE: DNA

; ORGANISM: not required under old rule

US-08-976-063E-21

Alignment Scores:

Pred. No.: 9,538-36 Length: 1518  
Score: 411.50 Match: 137  
Percent Similarity: 40.68% Conservative: 79  
Best Local Similarity: 25.80% Mismatches: 192  
Query Match: 13.08% Indels: 123  
DB: 4 Gaps: 19

US-09-758-269-6 (1-599) x US-08-976-063E-21 (1-1518)

Qy 118 AspProSerValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgArg 137

Db 1497 GACCG-----CAATTAGTAGGAACACTTCTCCCCCGTATAGAGCAGACTTGTTC 1444

Qy 138 AsnLeuProValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGly 157

Db 1443 GATCTAGAGTTGACGCGCAATCCCAAAATCAATAAATGGAACGCTTCTACCGTAATACG 1384

Qy 158 AlaAsnProLeuHisGluProValThrGlyHisPhePheAspGlyAspGlyMetVal 177

Db 1383 CCAGAGCTCAAGTATCCCAACAAATTCACACTTCATAGATGGAGATGGAAATGCC 1324

Qy 178 HisAlaValLysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsn 197

Db 1323 TCTGCTTCCACTTCGAAGATGTCATGTCGACTTCATCAGCTCGCTGGGTTAAACCGCT 1264

Qy 198 ArgPheValGlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeu 217

Db 1263 CGATTACGCGCGCAACGACTAGCGCAAAATCGCTATTTCGCGATGACAGAAACCCCTAT 1204

Qy 218 HisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaGlyIle 237

Db 1203 ACCGACGACACAGTGTAAAGGACTA----- 1177

Qy 238 ValAspProAlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArg 257

Db 1176 -----GACGCGACCGTTGCCAATACAGCATCATTAGCATTCACGCGCAAG 1132

Qy 258 LeuLeuAlaMetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAsp 277

Db 1131 GTCCTGGCGGTGAAGAGACGCGCTACCGTAC-----GAACTGGATCCTCGT---ACA 1081

Qy 278 LeuLysThrValGlyArgPheAspPheAspGlyGlnLeuGluSer---ThrMetIleAla 296

Db 1080 CITGAACTCGCGGACACTTCGACTACGACGCGCAAGTTACCGACCAACCCACACCGCC 1021

Qy 297 HisProLysValAspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSer 316

Db 1020 CATCCAAATATGACCGGAAACGGGTGACTTGTG----- 985  
Qy 3.17 LysProTyrLeuLysTyrPheArgPheSerProAspGlyThrLysSerProAspVal--- 335  
Db 984 -----TTCTTCGGTTCGGCAGCTAAGGCGGAGCAACTCCAGACATGGCC 940  
Qy 336 -----GluileGlnLeuAspGlnPro--- 342  
Db 939 TATTACATTTCGACACAGCAGCGCAAGGTGACACATGAACCTGGTTTGTAGCAGCCCTAT 880  
Qy 343 ---ThrMetMetHisAspPheAlaIleThrGluAsnPhaValValProAspGlnGln 361  
Db 879 GCGCATTCATGACAGCTTTCGCAATACCGAAATGTCATTTCCCAATATTATGCGG 820  
Qy 362 ValValPheLysLeuProGluMetileArgGlyGlySerProVal---ValTyrAspLys 380  
Db 819 GCCACCAACAGCGCTG---TCCCGCTCAAGCGGAAACAGCAATTTATATGTGGAGCG 763  
Qy 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLys 400  
Db 762 GAACTGGGCGACATACATTGGCGTACTC-----GCGCGCGCGCGGAGGCTGCTATTGCG 709  
Qy 401 TrpIleAspAlaProAspCysPheCysPheHisLeuTrpAsnAlaTrpGluGluProGlu 420  
Db 708 TGCTCAAGGACCGCGCTGCGTATTTCATGTTGTGAATGCTTGGAAAGTCGGAACC 649  
Qy 421 ThrAspGluValValIleGlySerCysMetThrPro-----ProAspSer--- 436  
Db 648 AAGATTATATGACCTTATGAAAGTGAATTCCTGCCGTTCCCTTCCCACTCAAA 589  
Qy 437 -----11e 437  
Db 588 AACCAACCTTCGCCCTCGAGAAAGCGGTGACACCGCTGACTCGTGGTGGAAATTCACCTC 529  
Qy 438 PheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLys 457  
Db 528 GATGACGAGCGGACGAGATCAAGCGAACCGGCTACACGATTCTTTGGGAAATGCCA 469  
Qy 458 ThrGlyGluSerThr----- 462  
Db 468 ATCATGATCTTCGTTCCCTCGATGCAATGCAACCGCTATGGCTTTATGGGGTGGACCAT 409  
Qy 463 ---ArgArgProIleIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMetVal 481  
Db 408 CCACGCAAAACCACTTCGCGATCAGCAGCGCGGAGAGATA----- 370  
Qy 482 AsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluProTrp 501  
Db 369 -----TTCGGGTCAACTCACTCGGC---ATCTGG 343  
Qy 502 ProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHisLeu 521  
Db 342 -----GACAAACCGGAGGTGACTACGACCTCTGGTAC 310  
Qy 522 TyrGlyAspAsnArgTyrGlyGluProLeuPheLeuPro---GlyGluGlyGlu 540  
Db 309 TCGGCGAAGCCTCGCGCGCCAGGAGCGGCTTCGTCCTCTAGAGTCCGACCGCCGCC 250  
Qy 541 GluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLysSerGluLeu 560  
Db 249 GAAGGTGACGGGTACTTCTGCTACCGGTGTTGTGCGCTCGATGATAAATCGCAGCATCTG 190  
Qy 561 GlnIleValAsnAlaValSerLeuGlu-----ValGluAlaThrValLysLeuProSer 578  
Db 189 GTAAATCTCGACACTCAAGACATCCAGTCTGGTCCGTCGCGTGGCAACCATCAAGTGCATTC 130  
Qy 579 ArgValProTyrGlyPheHisGlyThrPheIle 589  
Db 129 CGGTAAGGCGCGCTCTCCATGCTGCTGGTA 97

## RESULT 7

US-08-200-807-1

; Sequence 1, Application us/08200807

Patent No. 5573939  
GENERAL INFORMATION:  
APPLICANT: B Vik, Claes Olof, Eriksson, Ulf  
TITLE OF INVENTION: Isolated Protein Receptors, Antibodies Which  
TITLE OF INVENTION: bind thereto, Nucleic Acid Sequence Coding  
Patent No. 5573939  
TITLE OF INVENTION: Therefor, And Uses Thereof  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felfe & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/200,807  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/883,539  
FILING DATE: 15-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Hanson, No. 5573939man D.  
REGISTRATION NUMBER: 30,946  
REFERENCE/DOCKET NUMBER: LUD 280  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2629 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: no  
ANTI-SENSE: no  
US-08-200-807-1  
Alignment Scores:  
Pred. No.: 5,23e-13 Length: 2629  
Score: 211.00 Matches: 130  
Percent Similarity: 40.03% Conservative: 99  
Best Local Similarity: 22.73% Mismatches: 215  
Query Match: 6.70% Indels: 131  
DB: 26 Gaps:  
US-09-758-269-6 (1-599) x US-08-200-807-1 (1-2629)  
Qy 105 ValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSerValGlnIleAla 124  
Db 14 ATGTCAGCGCAAGTGTGACATCCA-----GCT 40  
Qy 125 GlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuPro----- 140  
Db 41 GGTGTTACAAGAACTGTTTGAACCTGTGAGGAACTATCTCTACCGCTCACAGCCCAT 100  
Qy 141 ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnPro 160  
Db 101 GTTACAGCGAGGATCCCTCTCGCTAACCGGAGTCTCTCTTCATGTGGCCAGGACTC 160  
Qy 161 LeuHis-----GluProValThrGlyHisHisPheAspGlyAspGlyMetVal 177  
Db 161 TTTGAGGTGGATCGGAACCATTT-----TACCACCTGTTTGTATGGCAAGCCCTCTTA 214  
Qy 178 HisAlaValLysPheGluHisGlySerAlaCysArgPheThrGlnThrAsn 197

Db 215 CACAAGTTTGACTTTAAAGAGGACATGTCACATACACAGAGGTTTCATCCGACGTGAT 274  
Qy ArgPheValGlnGlu-----ArgGlnLeuGlyArgProVal 209  
Db 275 GCTTACGTACGGGCAATGACTGAGAAAGGATGCTATACAGAAATTTGGACCTGTGCT 334  
Qy PheProLysAlaIleGlyLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPhe 229  
Db 335 TTCCAGATCCCTGCAAGAAATA-----TTTCCAGGTTTTTTTCTTAC 379  
Qy TyrAlaArgAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsnAla 249  
Db 380 TTC-----CGAGGAGTGGAGGTTTACTGACAAT 406  
Qy GlyLeuValTyrPhe-----AsnGlyArgLeuLeuAlaMetSerGluAspLeu 266  
Db 407 GCCCTTGTTAA-TATCTACCCAGTGGGGAAGATTACTATCCCTGCGACAGACCAACTT 465  
Qy ProTyr-GlnValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPh 286  
Db 466 CATTACAAGGTT-----AATCCTGAG---ACCTTGGAAACAATTAAAGCAGGTTGACCT 516  
Qy eAspGlyGlnLeuGlu---SerThrMetIleAlaHisProLysValAspProGluSer-- 304  
Db 517 TTGCAACTATGTCCTCAGTCAATGAGGACCACTGCTCACCCACATTTGAAATGATGGAC 576  
Qy -----GlyLeuLeuPheAlaLeuSerTyrAspValValse 316  
Db 577 TGTGTACAACATTTGTAATGCTTTGGGAAAAATTTTCAATTGCTCACTAATTGTA 636  
Qy rLysProTyrLeuLysTyrPheArgPheSerProAspGlyThrLysSerProAspValG1 336  
Db 637 GATCCACACATACACAGCAGACAGAGATCCA-----ATAAGCAAGTCAGAGATCGT 690  
Qy -----GlnProThrMetMetHisAspPheAlaIleTh 351  
Db 691 TGTACAATTCCTCCCTGAGTGACCGATTCAAGCCATCTTACGTCATAGTTTGGTTGAC 750  
Qy rGluAsnPheValValValProAspGlnGlnValPheLysLeuProGluMetIleAr 371  
Db 751 TCCCAACTATATTGTTTGTGGAGACACCACTGCAAAATTAATCTGTTCAAGTTTCITTC 810  
Qy gGlyGlySerProValValTyrAspLysAsnLysValAlaArgPhe----- 386  
Db 811 TTCATGGAGT-----CITTTGGGAGCCATTTACATGGATTGTTTGAATCAATGAAC 864  
Qy -----GlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLysTr 401  
Db 865 CATGGGGGTTTGGCTTCATATTGCTGCACAAAAAAGAAAAAGTATATCAATAATAATA 924  
Qy pIleAspAlaProAspCysPheHisLeuTrpAsnAlaTrpGluGluProGluTh 421  
Db 925 CAGGACCTCTCT----TTTAACTCTTTCATCATCATCACTATGAGACCATGATT 981  
Qy rAspGluValValValIleGlySerCysMetThrProAspSerIlePheAsn----- 439  
Db 982 T-----CTGATTGTGGATCTCTGTTGCTGGAAAGATTGAATTTGTTTATAATTATT 1035  
Qy -----GluSerAspGluAsnLeuLysSerValLeuSerGluIleAr 453  
Db 1036 ATATTAGCAATTTAGTGAACATGGGAGAGGTTGAAAAA---ATGCCAGAAAGCT 1093  
Qy gLeuAsnLeuLysThrGlyGluSer-----ThrArgAr 464  
Db 1094 CCTCAGCTGAATTTAGGAGATACGTACTCTCTTTGAATATTGACAAAGGCTGACACAGGC 1153  
Qy gProIleIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMetValAsnArgAs 484  
Db 1154 AAGAAATTAGTCACATCCCAACACAACTGCCATCGCAATTCGTGCGAGTACGAGACC 1213  
Qy nMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluProTrpProLysVa 504  
Db 1214 ATCT---GGCTGGAACCTGAGGTTCTCTTTTTCAGGGCCCTCGCAAGCAATTTGAGTTTCT 1270

Qy 504 lSerGlyPheAlaLys-----ValAspLeuThr-- 513  
Db 1271 CAAATCAATTACCAGAAAGTGTGGAAACCTTACACATATGATGACTTGGCTTG 1330  
Qy 514 -----ThrGlyGluValLysLysHisLeuTy 522  
Db 1331 AATCACTTTGTTTCCAGACAGGCTCTGTAAGCTGGAACGTCAAAACATAAAGAAACCTGGGTA 1390  
Qy 522 rGlyAsp-----AsnArgTyrGlyGlyGluProLeuPheLeuPro---GlyGluGlyG1 539  
Db 1391 TGGCAA-GAGCCTGATTATACCCCTCAGACCTATCTTTGTTTCTCACCAGATGCTT 1449  
Qy yGluGluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLys----- 557  
Db 1450 CGAGGAAGATGACGGTGTAGTTCTGAGTGTGTGTGAGCCCTGGGGCAGGACAAAAGCC 1509  
Qy 558 -SerGluLeuGlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuPr 577  
Db 1510 TGTCTATCTTCGATTCTGAAATGCCAAGACTTGAGTGAAGTTGCCAGGGCTGAAGTGA 1569  
Qy 577 oSerArgValProTyrGlyPheHisGlyThrPhe 588  
Db 1570 GATTAACATCCCGTCACCTTTTCATGGACTGTTTC 1603

RESULT 8  
US-08-488-305A-1  
; Sequence 1, Application US/08488305A  
; Patent No. 5679772  
; GENERAL INFORMATION:  
; APPLICANT: B vik, Claes Olof; Eriksson, Ulf; Peterson, Per A.  
; TITLE OF INVENTION: Isolated Protein Receptors, Antibodies Which  
; TITLE OF INVENTION: bind Thereto, Nucleic Acid Sequence Coding  
; Patent No. 5679772  
; TITLE OF INVENTION: Therefor, And Uses Thereof  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Felfe & Lynch  
; STREET: 805 Third Avenue  
; CITY: New York City  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 kb storage  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/488,305A  
; FILING DATE: 7-JUNE-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kohli, Vineet  
; REGISTRATION NUMBER: 37,003  
; REFERENCE/DOCKET NUMBER: LUD 5280.3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 688-9200  
; TELEFAX: (212) 838-3884  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2629 bases  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; HYPOTHETICAL: no  
; ANTI-SENSE: no  
US-08-488-305A-1

Alignment Scores: 5.23e-13 Length: 2629  
Pred. No.: 211.00 Matches: 130  
Score:





## Alignment Scores:

Pred. No.: 3,3e-13 Length: 1724  
Score: 210.00 Matches: 113  
Percent Similarity: 37.08% Conservatives: 98  
Best Local Similarity: 19.86% Mismatches: 218  
Query Match: 6.67% Indels: 140  
DB: 3 Gaps: 25

US-09-758-269-6 (1-599) x US-09-385-259-1 (1-1724)

QY 120 SerValGlnIle-----AlaGlyAsnPheAlaProValAsnGluGlnProValArg 136  
DB 31 TCCATCCAGTGGAGCATCCCGCGCGGTACAGAAAGCTGTTTGAACCGTGGAAAGAG 90  
QY 137 ArgAsnLeuPro-----ValValGlyLysLeuProAspSerIleLysGlyVal 152  
DB 91 CTGTCGTGCGCGCTCACCGCCACGTCAGAGGATCCCGCTCTGCTCAGCGGCGAGT 150  
QY 153 TyrValArgAsnGlyAlaAsnProLeuHis-----GluProValThrGlyHisHis 169  
DB 151 CTCCTCCGATCGGACCGCGGCTCTCGAGGTGGATCTGAACCATTT-----TACCAC 204  
QY 170 PhePheAspGlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerTyr 189  
DB 205 CTGTTTGACGACAGCCCTTCTGCACAACTTCGACTTTAAAGAAAGACACGTCACCTAT 264  
QY 190 AlaCysArgPheThrGlnThrAsnArgPheValGlnGlu----- 202  
DB 265 CACAGAAGGTTTACCGACCGATGCTTACGTCGCGGCAATGACCGAGAAAGAGATCGTC 324  
QY 203 ---ArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHisThr 221  
DB 325 ATAACGGAATTGGCACCTGTCGCTCCAGATCCCTGCAAGATATA----- 372  
QY 222 GlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAspProIle 241  
DB 373 ---TTTCCAGGTTTTTTCTTACTTC----- 396  
QY 242 HisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAlaMet 261  
DB 397 CGAGGAGTGGAGTCACTGCAATATGCCCTTGTT-----AAGCTTACCCAGTA 444  
QY 262 SerGluAsp-----AspLeuProTyrGlnValGlnIleThrProAsnGly 276  
DB 445 GGGGAAGATTACTACGCTGCGCAGGAGACCAACTTACAAAGATTAACTCTGAG--- 501  
QY 277 AspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeuGlu---SerThrMetIle 295  
DB 502 ACCCTGGACAAATTAAAGCAGGTGTACTCTGCAACTACGTCCTCTGTCATGGAGCCACC 561  
QY 296 AlaHisProLysValAspProGluSer-----GlyGlu 306  
DB 562 GCTCACCCCCACATTGAAATGATGGGACTGTGTTACACATTGGTAATTGCTTTGGAAA 621  
QY 307 LeuPheAlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSer 326  
DB 622 AATTTTTCGATTCCTTACAAATATTGTAAGATCCCTCCACTCAAAGCAGACAAAGAAAT 681  
QY 327 ProAspGlyThrLysSerProPheValGluIleGlnLeuAsp-----Gln 341  
DB 682 CCA-----ATAGCAGTCCGAGTGGTCTGTCACATTTCCCTGCGAGCCGCAATCAAG 735  
QY 342 ProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAspGlnGln 361  
DB 736 CCATCGTACGTCCTATGTTTGGTTTGGTCTCCCACTATATGTTTGTGAGAGCGCA 795  
QY 362 ValValPheLysLeuProGluMetIleArgGlyGlySer----- 374  
DB 796 GTCAAAATTAACCTGCTCAAGTTCCTTTCTCGTGGAGTCTTTGGGAGCCCACTACATG 855  
QY 375 -----ProValValTyrAspLysAsnLysValAla 384  
DB 856 CATTTGTTTTCAGTCCAAATGAACCATCGGGGTTTGGCTTCACATCGCTGACAAAAAGA 915

QY 385 ArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLysTyrIleAspAla 404  
DB 916 AAAAAGTATCTCAATAAATAAGTACAGGACCTCTCTCTTAATCTC----- 960  
QY 405 ProAspCysPheCysPheHisLeuThrPheAsnAlaTyrGluGluProGluThrAspGluVal 424  
DB 961 -----TTCCATCATATCATATCTTACGAGACAAATGAGTTT-----CTG 999  
QY 425 ValValIleGlySerCysMetThrProProAspSerIlePheAsnGluSerAsp---Glu 443  
DB 1000 ATTGTGGATCTCTGCTGCTGGAAGGATTGAATTCGTCTACAAATTACTTGTATTAGCC 1059  
QY 444 AsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyLysSerThrArg 463  
DB 1060 AATTTTACGTGAGAACTCCGAGAGAGGTGAAA-----AAAAATCCCAAGAGGCTCG 1110  
QY 464 ArgProIleIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMetValAsnArg 483  
DB 1111 CAGCCTGAAGTAGGAGATCCGCTGCTTCTCTCAGGCGCTCGTCAAGCCTTTGAGTTTCTCAAATC 1170  
QY 484 AsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeu----- 497  
DB 1171 AACCTAGTCACTCTCCCAACACGAGCGCCACTGCAACTCTGGCAGCCGACGAGCCATC 1230  
QY 498 ---AlaGluProTyrProLysValSerGly-----PheAlaLysVal 510  
DB 1231 TGCTGGAACCTGAGTCTCTCTCAGGCGCTCGTCAAGCCTTTGAGTTTCTCAAATC 1290  
QY 511 AspLeuThrThr-----GlyGluValLysLysHisLeuTyrGly----- 523  
DB 1291 AACTATCAGAAGTATGGCGGAAGCCTTACACGTACGCGTATGGAATGGCTTGGCTGAATCAC 1350  
QY 523 ----- 523  
DB 1351 TTGTTCCGAGACAGGCTCTGCAAGCTGAAGCTCAAGCTCAAGAAACGTTGGTATGCAA 1410  
QY 524 ---AspAsnArgTyrGlyGlyGluProLeuPheLeuPro---GlyGluGlyGlyGluGlu 541  
DB 1411 GAGCCGCACTCATACCCATCAGAACCCATCTTTGTTTCTCACCAGATGCTTTGGAAGAA 1470  
QY 542 AspGluGlyTyrIleLeuCysPheValHisAspGluLysThrLys-----SerGlu 559  
DB 1471 GATGATGGTGTAGTTCTGAGTGTGGTGTGAGCCCTGGGCGAGCAAAAGCTGCTTAT 1530  
QY 560 LeuGlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArg 579  
DB 1531 CTCTGATTTCTGAATGCCAAGGATTGAGTGAAGTTGCCAGGCTGAAGTGGAGATTAAAC 1590  
QY 580 ValProTyrGlyPheHisGlyThrPhe 598  
DB 1591 ATCCCTGTCACCTTTCATGAGCTGTTTC 1617

## RESULT 10

US-09-645-370-1  
; Sequence 1, Application US/09645370  
; Patent No. 6428958  
; GENERAL INFORMATION:  
; APPLICANT: Aguirre, Gustavo D.  
; APPLICANT: Acland, Gregory M.  
; APPLICANT: Ray, Kunal  
; TITLE OF INVENTION: IDENTIFICATION OF CONGENITAL STATIONARY NIGHT BLINDNESS  
; FILE REFERENCE: 19603/2481  
; CURRENT APPLICATION NUMBER: US/09/645,370  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 09/385,259  
; PRIOR FILING DATE: 1999-08-20  
; PRIOR APPLICATION NUMBER: 60/103,219  
; PRIOR FILING DATE: 1998-10-06  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1

; LENGTH: 1724  
 ; TYPE: DNA  
 ; ORGANISM: Canis familiaris  
 US-09-645-370-1

## Alignment Scores:

Pred. No.: 3,3e-13 Length: 1724  
 Score: 210.00 Matches: 113  
 Percent Similarity: 37.08% Conservatives: 98  
 Best Local Similarity: 19.86% Mismatches: 218  
 Query Match: 6.67% Indels: 140  
 DB: 4 Gaps: 25

US-09-758-269-6 (1-599) x US-09-645-370-1 (1-1724)

QY 120 SerValGlnIle-----AlaGlyAsnPheAlaProValAsnGluGlnProValArg 136  
 Db 31 TCATCCAGTGGAGCATCCCGCGCGTTTCAAGAGCTGTTTGAACCGTGAAGAG 90  
 QY 137 ArgAsnLeuPro-----ValValGlyLysLeuProAspSerIleIleGlyVal 152  
 Db 91 CTGTCCTCGCGTCAACCGCCACGCGTGACAGGAGGATCCCGCTCTGGCTCAGGGCAGT 150  
 QY 153 TyrValArgAsnGlyAlaAsnProLeuHis-----GluProValThrGlyHisHis 169  
 Db 151 CTCCTCCGATGCGGACCGCGGCTCTTCGAGGTGGATCTGAACCATTT-----TACCAC 204  
 QY 170 PhePheAspGlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerTyr 189  
 Db 205 CTGTTTGACGGAACAGCCCTCTGCACAGTTCGACTTTAAAGAGGACACGCTCACTAT 264  
 QY 190 AlaCysArgPheThrGlnThrAsnArgPheValGlnGlu----- 202  
 Db 265 CACAGAAGGTTCATCCGACCGATGCTTACGTCGGGCAATGACGAGAAAGGATCGTC 324  
 QY 203 ---ArgGlnLeuGlyArgProValPheProLysAlaIleGlyLysHisGlyHisThr 221  
 Db 325 ATACGGAATTTGGCAGCTGCTGCTCCAGATCCCTCGAAGATATA----- 372  
 QY 222 GlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaGlyIleValAspProAla 241  
 Db 373 ---TTTTCCAGTTTTTTTCTTCTTCTT----- 396  
 QY 242 HisGlyThrGlyValAlaAlaAsnAlaGlyLeuValTyr-PheAsnGlyArgLeuLeuAlaMet 261  
 Db 397 CGAGAGTGAGGTCACTGACATGACCTGTT-----AACGCTCAACCGATA 444  
 QY 262 SerGluAsp-----AspLeuProTyrGlnValGlnIleThrProAsnGly 276  
 Db 445 GGGGAAGATTACTACGCTGTCACGAGACCAACTTATTACAAAGATTATCTCTGAG--- 501  
 QY 277 AspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeuGlu---SerThrMetIle 295  
 Db 502 ACCCTGGAGCAATTAAGCGTTGATCTCTGCACTACGCTCTGTCATGAGGCCACC 561  
 QY 296 AlaHisProLysValAspProGluSer-----GlyGlu 306  
 Db 562 GCTCACCCCCACATTGAAATGATGGGACTGTTTACAACTGTAATGCTTTGGGAAA 621  
 QY 307 LeuPheAlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSer 326  
 Db 622 AATTTTTCGATTGCTCAATATTGTAAGATATCCCTCACTCCAAAGACAGACAGGAGAT 681  
 QY 327 ProAspGlyThrLysSerProAspValGluIleGlnLeuAsp-----Gln 341  
 Db 682 CCA-----ATAAGCAAGTCCGAGGTGCTCGTACAAATTCCTCCGACGACCGATTCAAG 735  
 QY 342 ProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAspGlnGln 361  
 Db 736 CCATCGTACGTCATAGTTTGGTTGATCCCACTATATTGTTTGTGGAGACCCCA 795  
 QY 362 ValValPheLysLeuProGluMetIleArgGlyGlySer----- 374

Db 796 GTCAAAATTAACCTGCTCAAGTTCCTTCTCTGAGAGTCTTTGGGAGCCCACTACATG 855  
 QY 375 -----ProValValTyrAspLysAsnLysValAla 384  
 Db 856 GATTGTTTGAAGTCCAAATGAACCATGGGGTTCCTTCATCATCGTCACAAAAAAGA 915  
 QY 385 ArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLysTrpIleAspAla 404  
 Db 916 AAAAAGTATCTCAATAAAGTACAGGACCTCTCTCTTAATCTC----- 960  
 QY 405 ProAspCysPheCysPheHisLeuTyrAsnAlaTyrGluGluProGluThrAspGluVal 424  
 Db 961 -----TTCCATCATATCAATCTACGAACAAATGAGTTT-----CTG 999  
 QY 425 ValValIleGlySerCysMetThrProProAspSerIlePheAsnGluSerAsp---Glu 443  
 Db 1000 ATTTGATGATCTGCTGCTGGAAGGATTTGAATTCGTCTACAATTTACTTGTATTAGGC 1059  
 QY 444 AsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGluSerThrArg 463  
 Db 1060 AATTTAGTTCGAGACTGGGAGAGGTGA---AAAAATCCCAAGAAAGGCTCCG 1110  
 QY 464 ArgProIleIleSerAsnGluAspGlnValAsnLeuGluAlaGlyMetValAsnArg 483  
 Db 1111 CAGCTCAAGTTAGGAGATCCGTGCTTCTTGAATATGACAAAGCCGACACAGGCAAG 1170  
 QY 484 AsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeu----- 497  
 Db 1171 AACTAGTCACTCCCTCCCAACAGCGGCCACTGCAACTCTGCGAGCGACGAGACCATC 1230  
 QY 498 ---AlaGluProTyrProLysValSerGly-----PheAlaLysVal 510  
 Db 1231 TGGCTGGAACCTGAGGTTCCTTCTCAGGGCCTCGTCAAGCCTTTCAGTTTCTCCTCAATC 1290  
 QY 511 AspLeuThrThr-----GlyGluValLysLysHisLeuTyrGly 523  
 Db 1291 AACTATCAGAAGTATGCGGGAAGCCCTTACAGTACCGGTATGCACTTGGCTTGAATCAC 1350  
 QY 523 ----- 523  
 Db 1351 TTCGTTCCGACAGGCTCTGCAAGCTCAAGCTAAAGAAACGTTGGTATGCAA 1410  
 QY 524 ---AspAsnArgTyrGlyGlyGluProLeuPheLeuPro---GlyGluGlyGlyGluGlu 541  
 Db 1411 GAGCCGACTCATACCATCCATCAGAACCCATCTTGTCTTCCACCCAGATGCTTGGAAAGAA 1470  
 QY 542 AspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLys-----SerGlu 559  
 Db 1471 GATGATGTTGATGTTCTGAGTGTGTGTGTGAGCCCTGCGGACGACAAAGCCCTGCTAT 1530  
 QY 560 LeuGlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArg 579  
 Db 1531 CTTCTGATTCGTAATGCAAGGATTTGAGTGAAGTTGCGCAGGGCTGAAGTGAAGATTAAAC 1590  
 QY 580 ValProTyrGlyPheHisGlyThrPhe 588  
 Db 1591 ATCCCTGTCACTTTCATGGACTGTTCT 1617

## RESULT 11

US-08-311-731A-140  
 ; Sequence 140, Application US/08311731A  
 ; Patent No. 6583266  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SMITH, DOUGLAS  
 ; APPLICANT: MAO, JEN-I  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
 ; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR  
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
 ; NUMBER OF SEQUENCES: 411  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
 ; STREET: 600 ATLANTIC AVENUE  
 ; CITY: BOSTON

STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/311,731A  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: GATES, EDWARD R.  
REGISTRATION NUMBER: 31,616  
REFERENCE/DOCKET NUMBER: C0044/7125  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/720-3500  
TELEFAX: 617/720-2441  
INFORMATION FOR SEQ ID NO: 140:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 36063 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Mycobacterium leprae  
US-08-311-731A-140  
Alignment Scores:  
Pred. No.: 2,74e-08 Length: 36063  
Score: 186.00 Matches: 133  
Percent Similarity: 38.61% Conservative: 72  
Best Local Similarity: 25.05% Mismatches: 209  
Query Match: 5.90% Indels: 121  
Gaps: 23  
US-09-758-269-6 (1-599) x US-08-311-731A-140 (1-36063)  
QY 110 LeuHisProLeuProLeuThraAlaAspProSerValGlnileAlaGlyAsnPhaAlaPro 129  
Db 22576 TTGTGACGACATATAGAGGGATCAGACATGATGTCGAATGTTCAGCAATGCTCTCT 22635  
QY 130 Val-----AsnGlnProValArgAsnLeuPro----- 140  
Db 22636 ACCCTGTCGGAAGACGACGACCTCGCGCGGATTGTCCCGTGGCGACCGACGACGATC 22695  
QY 141 -----ValValGlyLysLeuProAspSerIleIysGly 151  
Db 22696 GAATGGGACGCAACGACCTCAACGCTTTAGCGGAGAAATACCCAGCTACCTAAAGCG 22755  
QY 152 ValTyrValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHisHisPhePhe 171  
Db 22756 TACCTACCGACGCAACACATAAAACCACTGACCCGGAGTTTCACAGTGTATCAACCCGTT 22815  
QY 172 AspGlyAspGlyMetValHisAlaValIysPheGluHisGlySerAlaSerTyrAlaCys 191  
Db 22816 GATGAAGACGGCATGCTGACATGCTCGGCTCCCGATGAAAGCCCTTTTTCGCAAC 22875  
QY 192 ArgPheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPhePro 211  
Db 22876 ACCTTTATACGACACGACGATTTTGTCCGAGACATCGCAGGCGGCCCTGTGGCCC 22935  
QY 212 LysAlaIleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyrAla 231  
Db 22936 GG-TTG-----GGCAGAACCGGT----- 22952  
QY 232 ArgAlaAlaAlaGlyIleValAspProAla---HisGlyThrGlyValAlaAsnAlaGly 250  
Db 22953 -----GCATATATGACCAAGCGTGAACAGGCGCTGGGGATCCGCAACCGAGATG 23000

251 LeuValTyrPheAsn-----GlyArgLeuLeuAlaMetSerGlu----- 263  
23001 AAGGACGGGTTGAACACCGAOCCTCACGTCATCGAGGATCGCGCTACCGAGTTTCTAC 23060  
264 -----AspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThrVal 281  
23061 CATTGACGAGTCTG--TACCACATAGAACCTATCCC-----GCCAATACGCGA 23108  
282 GlyArgPheAspPheAspGlyGlnLeuGluSerThr---MetIle-AlaHisProLysVa 300  
23109 GGCAAGAGATCGGGGACGGCGTTTACAGTTGATGGAGCGGTGGCGACATCTCAAACT 23168  
300 lAspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSer-LysProTyrL 320  
23169 GGCAACACGACCGGAGAACTG-----CTGTTCTTTAATACACAGCAGGACCGGTACA 23222  
320 euLysTyrPheArgPheSerProAspGlyThrLys---SerProAspValGlnIleGlnL 339  
23223 CATCGGCTATGCCCTCAACCAACCAAGCAACGAACTGACACACACAGTCGACATTCGCG 23282  
339 euAspGlnProThrMetHisAspPheAlaIleThrGluAsnPheValValValProA 359  
23283 TGTGTGGTTCGGGTTGGCAGATGCGATGGCGGTTCCACGAAACTACACGATCTCAATA 23342  
359 spGlnGlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyrA 379  
23343 ATTG-CCCATTGTTCTGAAG-----TCCAAGACTGCTCAAG 23377  
379 spLysAsnLysValAlaArgPhe-----GlyIleLeuAspLysTyrAlaGluAspSerS 397  
23378 AACATGTGTGTACACGATTTCTATCGTACATGCTGTCTCGTTCGGGTGTT-TCNA 23436  
397 er-----AsnIleLysTrpIleAspAlaProAspCysPheCysPheHisLeuTrpAsnA 415  
23437 GCACCGGCGACATCGGTGATTCGAGGGGGAATATATATTCTTACCGCATTTCTCAACG 23496  
415 laTrpGluGluProGluThrAspGluValValIleGlySerCysMetThrProProA 435  
23497 CCTACGAA-----AAGGTGACGAGATCATG-----CGCG 23526  
435 spSerIlePheAsnGluSerAspGluAsnLeu-----LysSerValL 449  
23527 ATAGATTCTCGAAGACTAGCCACACACCTTTTCACCGCGCGCCACTGACCAAGCCGG-C 23585  
449 euSerGluIleArgLeuAsnLeuLysThrGlyLysThrArgArgProIleIleSera 469  
23586 TGCACAGCTGGCGATTACCCCTAGTCACCGGTCGTGTAAACAGAGTAACATG----- 23638  
469 snGluAspGlnGlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgL 489  
23639 -----CCGGAATCTATCACCGAGTTCGGGATGATCATGCTGATGCTCCAGCA 23690  
489 yThrLysPheAlaTyrLeuAlaLeuAlaGluProTyrProLysValSerGly----- 506  
23691 AGTATCACTACTCTATACCTATGCGGCC-----ACCGGCCAATCGGCTAGTTCT 23741  
507 -----PheAlaLysValAspLeuThrThrGlyGluVal-LysLysHisLeuTyr 522  
23742 TCTTACAGCTATTGGCTAAACACCTTCTGCTACCAACACCTAGAGCGCTACTCATTC 23801  
523 GlyAspAsnArgTyrGlyGlyGluProLeuPheLeuProGlyGlyGlyGlu---Glu 541  
23802 GCGATGGCATCTACGAAAGTGAGAGGCGGTGGGTCCGCGGTGGGTAGTACCAGGTGAA 23861  
542 AspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLys----- 557  
23862 GATGACGGCTAACTGGTAAACCCCTACCCACCGACATGAATGCCGATAGTCTTAATTATAT 23921  
558 -----SerGluLeuGlnIleValAsn 564  
23922 GGCGTTTCGAAGCAGCCAGGATCACCAGACAGTCCAGGTGAGTAAACTTCACTGCCGAAA 23981

QY 565 AlaValSerLeuGluValGluAlaThr 573  
 Db 23982 CGTATTTCCAGCAGCGCGCTTCCACC 24008

RESULT 12  
 US-09-252-991A-7056/c  
 ; Sequence 7056, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252.991A  
 ; PRIORITY FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 7056  
 ; LENGTH: 4242  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-7056

Alignment Scores:  
 Pred. No.: 0.00236 Length: 4242  
 Score: 127.50 Matches: 104  
 Percent Similarity: 35.96% Conservative: 65  
 Best Local Similarity: 22.13% Mismatches: 174  
 Query Match: 4.05% Indels: 129  
 DB: 4 Gaps: 21

US-09-758-269-6 (1-599) x US-09-252-991A-7056 (1-4242)

QY 23 ProLeuSerSerGlnSerSerAspLeuSerTyrCysSerSerLeuProMetAlaSer 42  
 Db 2486 CATGCACCTGCGCGGTGATCGAGTCTTGGCATGC-----GCGCGC 2442  
 QY 43 ArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProProAlaLeuHisPhe 62  
 Db 2441 AGGATTGCGAACTGCACCTTCTTATCGATCGACTCGAGCTGCCAGCGCGCTGTGG 2382  
 QY 63 ProLysGlnSerSerAsnSerProAlaIleValLysProLysAlaLysGlnSerAsn 82  
 Db 2381 CCCCGTTGTGTGCG--GCGCCCGGTGCTGCTCGGCGCCAGGCCAGTGGCGCGG 2325  
 QY 83 ThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAlaAlaGluGly 102  
 Db 2324 AGGAAA-----TCTGCAG-CTGATCCGCGCGCGAGGCG 2293  
 QY 103 PheLeuValSerHisGlnLysLeuHisProLeuProLysThrAlaAspProSerValGln 122  
 Db 2292 -----GTGAG-CATCTCGCTTACCCCGAGCTACGG----- 2261  
 QY 123 IleAlaGlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuProValVal 142  
 Db 2260 -----CAGCCAGCTCGC-----CCAGTGGCT 2240  
 QY 143 GlyLysLeuProAspSerIleLysGlyValTyrValArgAsn-GlyAlaAsnProLeuHi 162  
 Db 2239 GGAGAGCCAGGCGCGGAGTGGCGGTGCGATGCGATGCATCCCGCGGC----- 2191  
 QY 162 sGluProValThrGlyHisHis-----Ph 170  
 Db 2190 -GAAGCGCTACCGCGGAGCAGCACCTGCACGATTGCCAGGCTTCGCCCGCGCGT 2132  
 QY 170 sPheAspGlyAspGlyMetValHisAlaVal----- 180  
 Db 2131 CTTCAATGCTACGAGCCGAGACGCGGTGTCATCGCGCTGCGCTGCTGCTGCGA.2072  
 QY 181 -LysPheGluHisGlySerAlaSerTyrAlaCysA-gPheThrGlnThrAsnArgPheVa 200

Db 2071 ACCACTGGAGGAGGTGCGCGCAGCGTCCGATCGGACGCGTGTGCGCGCGGTGCG 2012  
 QY 200 lGlnGluArgGlnLeuGlyArgProValPheProLys---AlaIleGlyGluLeuHis-- 218  
 Db 2011 CTACATCTCGATGCGGACCTGGCCCTGGTCCGCGAGCGCGACCGGAACTCTAGT 1952  
 QY 219 -GlyHisThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGly---- 236  
 Db 1951 CGGCGCGCGCGCTGGCGCGC-----GGTACCATCAGCGTCCGCGCGCTCAGCGCGCA 1898  
 QY 237 -----IleValAspProAlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPh 254  
 Db 1897 GCGCTTCGTCGCGCATCCTTC-----GCTCCGAGGCGCGCGCTGTACCG 1850  
 QY 254 eAsnGlyArgLeuLeuAlaMetSerGluAspLeuProTyrGlnValGlnIleThrPr 274  
 Db 1849 CACCGCGCGACCTG-----GTGCGCTGTGCGA 1823  
 QY 274 oAsnGlyAspLeuLysThrValGlyArgPheAspPheAsp----- 287  
 Db 1822 CAACGCGCAGGTGGAATATGCGCGCATGCACCGAGTGAAGATCCGTGGCTTCG 1763  
 QY 288 -----GlyGlnLeuGluSerThrMetIleAlaHisProLysValAspProGluSe 304  
 Db 1762 CATCGAACTGGCGGAGATCGAGCGCGCTGCTGGAGCATCCGAGGT----- 1714  
 QY 304 rGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheAr 324  
 Db 1713 -CGCGAAGCGCTGCTGCTGCGCTGCAGCGCGCGAGCGGAGCGAGTGGCGGTATGT 1655  
 QY 324 gPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAspGlnProThrMe 344  
 Db 1654 CGCCAGCGGTGGCGGAGCAGCAGCAGCAGCGCGCGCTGCGCGAGCGCTGAA 1595  
 QY 344 tMetHisAspPheAlaIleThrGluAsnPheValValValProAspGlnGlnValValPh 364  
 Db 1594 AACGCTATCTCAAGACGAATGCGGACATACATGTCGCGCGCGCCACTCTGTGCTCG 1535  
 QY 364 eLysLeuProGluMetIleArgGlySerProValValTyrAspLysAsnLysValal 384  
 Db 1534 CAGCCTGCGC-----CTGAC 1520  
 QY 384 aArgPheGlyIleLeuAspLysTyrAla-----GluAspSerSerAsnIleLy 400  
 Db 1519 CGCAACGCGCAAGCTCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCTCAACCGGCA 1460  
 QY 400 sTrpIleAspAlaPro---AspCysPheCysPheHisLeuTyrAsnAlaTrpGluGluPr 419  
 Db 1459 GGCTACGAGCGCGCGCGCGCGCTGCGGAGCAGCAACTGGCGGCGGTCTGGCGCGAGGT 1400  
 QY 419 oGluThrAspGluValValValIleGlySerCysMetThrProAspSerIlePheAs 439  
 Db 1399 GCTGACGTCGAGCGGTAGGTCTCGC-----GACAACTTCTTCGA 1358  
 QY 439 nGluSerAspGluAsnLeuLysSerVal 448  
 Db 1357 ACTGCGCGCGATTCGATCTGTCGATC 1330

## RESULT 13

US-09-252-991A-6997  
 ; Sequence 6997, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252.991A  
 ; PRIORITY FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190

```
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6997
; LENGTH: 10023
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6997

Alignment Scores:
Pred. No.: 0.0102 Length: 10023
Score: 127.50 Matches: 104
Percent Similarity: 35.96% Conservative: 65
Best Local Similarity: 22.13% Mismatches: 174
Query Match: 4.05% Indels: 129
DB: 21

US-09-758-269-6 (1-599) x US-09-252-991A-6997 (1-10023)
QY 23 ProLeuSerSerSerGlnSerSerAspLeuSerTyrcysSerSerLeuProMetAlaSer 42
Db 6671 CCATGCACCTGCGCGCGGTGATCGATGCTTGGCATGC-----GCGCCG 6715
QY 43 ArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProAlaLeuHisPhe 62
Db 6716 AGGATTGCGAACTGCACCTTCTATTTCGATCAACTTCGACGCTGCCAGCGAGCGCTGCTGG 6775
QY 63 ProLysGlnSerSerAsnSerProAlaIleValLysProLysAlaLysGluSerAsn 82
Db 6776 CCCGTTCTGTGCG---GCGCCCGCTGTGCTGCGGCCCGAGCGAGTGGGGCGCG 6832
QY 83 ThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAlaAlaGluGly 102
Db 6833 AGGAAA-----TCTGCGAG-CTGATCCGCGCGCGAGGCG 6864
QY 103 PheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSerValGln 122
Db 6865 -----GTGAG-CATCCTCGGCTTACCCCGAGTACGG----- 6896
QY 123 IleAlaGlyAsnPheAlaProValAsnGlnProValArgArgAsnLeuProValVal 142
Db 6897 -----CAGCCAGCTCGC-----CCAGTGGCT 6917
QY 143 GlyLysLeuProAspSerIleLysGlyValTyValArgAsn-GlyAlaAsnProLeuHi 162
Db 6918 GGAGAGCCAGGCGCGCGAGTTCGCGGTGCGATGTCATCACCGCGCGC----- 6966
QY 162 sGluProValThrGlyHisHis-----ph 170
Db 6967 -GAAGCGCTGACCGCGAGCACTCAACGGATTCCGCCAGCGCTTCGCCCGCGCTGCTT 7025
QY 170 epheaspGlyaspGlyMetValHisAlaVal----- 180
Db 7026 CTTCAATGCTACGAGCCAGCCAGACGCGTGGTCAATCGCTGCGCTGCTGCTGCTGCTG 7085
QY 181 -LysPheGluHisGlySerAlaSerTyzAlaCysArgPheThrGlnThrAsnArgPheVa 200
Db 7086 AGCATGTGAGGAGGTGCGCGAGCGTCCGATCGCGAGCGTGTGCGCGCGCGGGTTCG 7145
QY 200 lGlnGluArgGlnLeuGlyArgProValPheProLys---AlaIleGlyGluLeuHis-- 218
Db 7146 CTACATCTGATGTCGCGACCTGCGCTGTCGCGAGCGCGCGCGCGAACTCTACGT 7205
QY 219 -GlyHisThrGlyIleAlaArgLeuMetLeuPheTyAlaArgAlaAlaAlaGly----- 236
Db 7206 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7259
QY 237 -----lleValAspProAlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyPh 254
Db 7260 GCGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7307
QY 254 eAsnGlyArgLeuLeuAlaMetSerGluAspAspLeuProTyrglnValGlnIleThrPr 274
Db 7308 CACCGCGGACCTG-----GTGCGCGCTGTGCGA 7334
```

```
274 oAsnGlyAspLeuLysThrValGlyArgPheAspPheAsp----- 287
Db 7335 CAACGCCAGGTGAATATGTGCGCGCATCGACACACAGGTGAAGATCCGTGGCTTCG 7394
QY 288 -----GlyGlnLeuGluSerThrMetIleAlaHisProLysValAspProGluSe 304
Db 7395 CATCGAATGCGCGAGATCGAGCGCGCTGCTGAGCATCGCAGGTT----- 7443
QY 304 rGlyGluLeuPheAlaLeuSerTyAspValValSerLysProTyLeuLysTyPheAr 324
Db 7444 -CGCGAAGCGGTGCTCTGCGGCTCGACAGCGCGAGCGAGCATGTTGGCGGTATGT 7502
QY 324 gPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAspGlnProThrWe 344
Db 7503 CGCCACGCGGTGCGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 7562
QY 344 tMethHisAspPheAlaIleThrGluAsnPheValValProAspGlnGlnValValPh 364
Db 7563 AAGCATCTCAAGCAGCAATTCGCGACTACATGTGCGCGCCACCTGCTGTCTCGC 7622
QY 364 eLysLeuProGluMetIleArgGlyGlySerProValValTyAspLysAsnLysValAl 384
Db 7623 CAGCCTGCGC-----CTGAC 7637
QY 384 aArgPheGlyIleLeuAspLysTyzAla-----GluAspSerSerAsnIleLy 400
Db 7638 CGCCACGCAAGCTCGACCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 7697
QY 400 sTrpIleAspAlaPro---AspCysPheCysPheHisIleuTrpAsnAlaItrpGluPr 419
Db 7698 GGCCTACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7757
QY 419 oGluThrAspGluValValValIleGlySerCysMetThrProAspSerIlePheAs 439
Db 7758 GCTGAACGTCGAGCGGTAGCTCGC-----GACAACTTCTTCGA 7799
QY 439 nGluSerAspGluAsnLeuLysSerVal 449
Db 7800 ACTGGCGCGGATTCGATCTCTCGATC 7827

RESULT 14
US-08-700-651-1
; Sequence 1, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; FILE REFERENCE: 480.19-4 (HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; EARLIER FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 1
; LENGTH: 5163
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-08-700-651-1

Alignment Scores:
Pred. No.: 0.0092 Length: 5163
Score: 123.50 Matches: 125
Percent Similarity: 34.60% Conservative: 85
Best Local Similarity: 20.59% Mismatches: 194
Query Match: 3.92% Indels: 203
DB: 35
```

US-09-758-269-6 (1-599) x US-08-700-651-1 (1-5163)

QY 5 ThrAlaThrAlaValSerGlyArgTrpLeuGlyGlyAsnHisThrGlnProLeu 24  
|||  
Db 2491 ACAATTCAGGTAATGTTTCAGGA-----ATT 2517

QY 25 SerSerSerGlnSerSerAspLeuSerTyrCysSerSerLeu-----ProMetAlaSer 42  
|||  
Db 2518 TCTGCAAGTGAATCAATTA---TTATCTCAGAAATCAGCTCTAATCGACCCAGCAACAAAT 2574

QY 43 ArgValThrArgLys-----LeuAsnValSerSerAlaLeuHisThrPro 58  
|||  
Db 2575 ATCGTTTGTGGGAATTTGGTGAATGTTGAACCCAGCAACAGGAGTGATGATTCAGGT 2634

QY 59 AlaLeuHisPheProLysGlnSerSerAsnSerProAlaIle-----ValVal 74  
|||  
Db 2635 TTTTGTAGGTCATCAGCAACAACTCAATTCCTCCCTGAGATTGAAGATGGTGTATTATT 2694

QY 75 LysProLysAlaLysGlnSerAsnThrLysGlnMetAsnLeu----- 88  
|||  
Db 2695 CCTCCAGAAAGTAGCAGCAGCAAAATGCTGATAAATTCAGTTATCTATCTCCCTCCAGCGTA 2754

QY 89 -----PheGlnArgAlaAlaAlaAlaLeuAsp 98  
|||  
Db 2755 CCAGAAATCAATTCAGAAAGGATCAGAAAGATTGATTCTTATTTCTGAATTTGATGAT 2814

QY 99 AlaAlaGluGlyPheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAsp 118  
|||  
Db 2815 ATTGAATCAGGTAGACTTATTTGGTCAAGTATCAAGAGACCAATCCAGGTTC 2868

QY 119 ProSerValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgAsn 138  
|||  
Db 2869 -----ATTGCTGTGTACTTGAACCCAAATAATGAAGACACCAACACAACTGAC 2916

QY 139 LeuProValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAla 158  
|||  
Db 2917 ---AGTGAATCTGTAAACCAATCGATCCACACAGGTCTG----- 2955

QY 159 AsnProLeuHisGluProValThrGlyHisPhePheAspGlyAspGlyMetValHis 178  
|||  
Db 2956 ---CCTTTCAATCCACCA---ACTGGTCAT-----TTGATTAAC 2988

QY 179 AlaValLysPheGluHisGlySerAlaSerTyr-----AlaCysArgPheThrGlnThr 196  
|||  
Db 2989 CCACAAATATATACCATGGATCTCTTCATTTCTGGTGCCATACAAATATGAGTTTCA 3048

QY 197 AsnArgPheValGlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGlu 216  
|||  
Db 3049 AATCGTATTAACTGATGATAATGTTTATGTTTACCAGTT-----GGTGAA 3093

QY 217 LeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaGly 236  
|||  
Db 3094 ATAACAGGTTTACCAAAAGGATCCAGGCTCAGATATCCATTTAACTCAACTCAGGTGAA 3153

QY 237 IleValAspProAlaHisGlyThrGlyValAlaAsn-----AlaGlyLeuValTyrPhe 254  
|||  
Db 3154 TTATTTGATTCATCAACAGGAAACCAATTAACATTTCTACTGCTGTGTTGTT----- 3207

QY 255 AsnGlyArgLeu-----LeuAlaMetSerGluAspLeuProTyrGlnValGlnIleThr 273  
|||  
Db 3208 AGTGGAACCACTGGCTTACCACCTATTGAAGATGAA-----PheAsp 287

QY 274 ProAsnGlyAspLeuLysThrValGlyArgPheAsp-----TTTGATCCATCAACTTCCCAATAGAT 3285

QY 3244 ---AATGGTAATTG----- 3285

QY 288 GlyGlnLeuGluSerThrMetIleAlaHisProLysValAspProGluSerGlyGluLeu 307  
|||  
Db 3286 GGTAAAT-----AACCAATATTGTTAACCCAGAAACCAACAGCACT 3324

QY 308 PheAlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerPro 327  
|||  
Db 3327 ----- 3327

Db 3325 GTCTCAGGATCAACTTCAGGTACTACAAACCA-----AAACCA 3363

QY 328 -----AspGlyThrLysSerProAspValGluIle-----GlnLeuAsp 340  
|||  
Db 3364 GGAATTCAGTCAATGGTGGAGGTGTTTACTGTGAGAGAAGCTAAAGATCAAGCCGAT 3423

QY 341 Gln-----ProThrMet 344  
|||  
Db 3424 AAGGGTAAGGATGATTAATTTGTTCCACCACTAATTTCTATCAATAAAGATCCAGTAACA 3483

QY 345 MetHisAspPheAlaIleThrGluAsnPheValValProAspGlnValValPhe 364  
|||  
Db 3484 AATACTCAGTACAGTAATACTACTGTGTAACATTAATTAACCCAGAAAGAAAGTT--- 3540

QY 365 LysLeuProGluMetIleArgGly-----GlySerProValVal 377  
|||  
Db 3541 ---ATTCCAGGTTCACTTCCAGGCTCTCTCAACTATCCATCATTTCAATATCTCCACAA 3597

QY 378 TyrAspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSer 397  
|||  
Db 3598 ACTGATGAG-----ATTACAGGAAGCCAGTGTGATCTGTACT 3636

QY 398 AsnIleLysTrpIleAspAlaProAspCysPheHisLeuTrpAsnAlaTrpGlu 417  
|||  
Db 3637 GGTTCGCATAT----- 3648

QY 418 GluProGluThrAspGluValValIleGlySerCysMetThrProProAspSerIle 437  
|||  
Db 3649 GATCCATCTCAGGTGAATTTATCGATCTCGCAACTAAATTTCAATTTCCAGATCAGTT 3708

QY 438 PheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLys 457  
|||  
Db 3709 GCAGGTGAT-----GAAATCTCTCACTGAAGTA---TTGAACATTACA 3747

QY 458 ThrGlyGluSerThrArgArgProIleLeuSerAsnGluAspGlnValAsnLeuGlu 477  
|||  
Db 3748 ACAGATCAAGTAAACAGGTTTCCCAAT-----GATCTTGAA 3783

QY 478 AlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeu 497  
|||  
Db 3784 ACTGGTCTT----- 3792

QY 498 AlaGluProTrpProLysValSerGlyPheAlaLysVal-----AspLeuThrThr 514  
|||  
Db 3793 -----CCAAGATATCCAGTATCAGAGCTCCCACTTCCAAATGGTACTTGGTTGAT 3846

QY 515 GlyGluValLysLysHisLeuTyrGlyAspAsnArgTyrGlyGlyGluProLeuPheLeu 534  
|||  
Db 3847 CCATCAATAAATAAACCAATTCAGGTTC-----CATTCGGA-----TTTATT 3891

QY 535 ProGlyGluGlyGlyGlu 541  
|||  
Db 3892 AATGGTACATCTCGAGAACAA 3912

RESULT 15  
US-08-928-361B-4  
; Sequence 4, Application US/08928361B  
; Patent No. 6071518  
; GENERAL INFORMATION:  
; APPLICANT: Petersen, Carolyn  
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,  
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS  
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM  
; TITLE OF INVENTION: SPECIES INFECTIONS  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: PETERS, VERNY, JONES & BIKSA  
; STREET: 385 Sherman Avenue, Suite 6  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306-1840  
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/928,361B  
FILING DATE: 12-SEP-1997

CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 60/026,062  
FILING DATE: 13-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Verry, Hana  
REGISTRATION NUMBER: 30,518  
REFERENCE/DOCKET NUMBER: 480.76-1 (HV)  
TELEPHONE: 650-324-1677  
TELEFAX: 650-324-1678

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5163 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-928-361B-4

Alignment Scores:  
Pred. No.: 5163  
Score: 123.50  
Percent Similarity: 34.60%  
Best Local Similarity: 20.59%  
Query Match: 3.92%  
Indels: 203  
Gaps: 35

US-09-758-269-6 (1-5995) x US-08-928-361B-4 (1-5163)

QY	5	ThrAlaThrAlaAlaValSerGlyArgTrpLeuGlyGlyAsnHisThrGlnProProLeu	24	...
DB	2491	ACAAATGAGGATTTGTTTCAGGA	...	ATT 2517
QY	25	SerSerGlnSerSerAspLeuSerTyrCysSerSerLeu	...	ProMetAlaSer 42
DB	2518	TCTCAAGTGAAGTCAATTA--TTATCTCAGAAATCAGCTCTAATCGACCCAGCAACAAT	2574	...
QY	43	ArgValThrArgLys	...	LeuAsnValSerSerAlaLeuHisThrProPro 58
DB	2575	ATGGTGTGGAGAAATTTGGTGAGTTGTTGAACCCAGCAAGGAGTGATGATTCACGGT	2634	...
QY	59	AlaLeuHisPheProLysGlnSerSerAsnSerProAlaIle	...	ValVal 74
DB	2635	TTTTTAGGTCCATCAGACAAACTCAATTCCTCCCTGAGATTGAAGATGGTGGTATTATT	2694	...
QY	75	LysProLysAlaLysGlnSerAsnThrLysGlnMetAsnLeu	...	...
DB	2695	CCTCCAGAGTAGCAGCAGCAAAATCTGATAAATTCAGATTATCTATTCCTCCAGCGTA	2754	...
QY	89	...	...	PheGlnArgAlaAlaAlaLeuAsp 98
DB	2755	CCAGAAATCAATTCAGAAAGGATCAGAAAGATTGATTCTATTCTCGAATTGATGATGAT	2814	...
QY	99	AlaAlaGluGlyPheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAsp	118	...
DB	2815	ATTGAGTCAGGTAGAGTATTATTTGGTCAAGTATCAAAAGAGACCAATCCAGGTCA	2868	...
QY	119	ProSerValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgArgAsn	138	...
DB	2869	...	...	...
QY	139	LeuProValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAla	158	...
DB	2917	---AGTGTAACTGGTAAACCAATTCGATCCAAACACACAGGTCTG	2955	...

QY	159	AsnProLeuHisGluProValThrGlyHisHisPhePheAspGlyAspGlyMetValHis	178	...
DB	2956	---CCTTTCAATCCACCA--ACTGGTCAT	...	TTGATTAAAC 2988
QY	179	AlaValLysPheGluHisGlySerAlaSerTyr	...	AlaCysArgPheThrGlnThr 196
DB	2989	CCAAACAATAATAATACCATGATTCTTCATTGCTGTGCATACAAATATGCAATTTCA	3048	...
QY	197	AsnArgPheValGlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGlu	216	...
DB	3049	AATGGTATTAGACTGATAATGTTTATGTTTACAGTT	...	GGTGAA 3093
QY	217	LeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaGly	236	...
DB	3094	ATAACAGGTTTACCAAGGATCCAGGCTCAGATATATTCATTTAACTCAACTACAGGTGAA	3153	...
QY	237	IleValAspProAlaHisGlyThrGlyValAlaAsn	...	AlaGlyLeuValTyrPhe 254
DB	3154	TTAGTTGATCCATCAACAGGAAGCCAAATTAACAATTTCTACTGCTGGTATTGTT	...	3207
QY	255	AsnGlyArgLeu---LeuAlaMetSerGluAspLeuProTyrGlnValGlnIleThr	273	...
DB	3208	AGTGGAAACCTGGCTTACCACCTATTGAAGATGAA	...	3243
QY	274	ProAsnGlyAspLeuLysThrValGlyArgPheAsp	...	PheAsp 287
DB	3244	---AATGGTAAATTG	...	TTGATCCATCACTAACTGTCATAGAT 3285
QY	288	GlyGlnLeuGluSerThrMetIleAlaHisProLysValAspProGluSerGlyGluLeu	307	...
DB	3286	CGTAAT	...	AACCAATTAGTTAACCCAGAAACCAACAGCACT 3324
QY	308	PheAlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerPro	327	...
DB	3325	GTCTCAGGATCAACTTCAGGTACTACAAACCA	...	AAACCA 3363
QY	328	---AspGlyThrLysSerProAspValGluIle	...	GlnLeuAsp 340
DB	3364	GGAATTCACAGTCAATGGTGGAGGTGTTGACCTGATGAAGAGCTAAAGATCAAGCCGAT	3423	...
QY	341	Gln	...	ProThrMet 344
DB	3424	AAGGTAAGGATGGATTAATTTGTCACCACTAATTTCTATCAATAAAGATCCAGTACA	3483	...
QY	345	MetHisAspPheAlaIleThrGluAsnPheValValProAspGlnGlnValValPhe	364	...
DB	3484	AATACTCAGTACAGTAATACTACTGTGTACATTATTAAACCCAGAAACAGAAAGTT	...	3540
QY	365	LysLeuProGluMetIleArgGly	...	GlySerProValVal 377
DB	3541	---ATTCAGGTTCACTTCCAGGCTCTCTCAACTATCCATCATTCATCACTCCACACAA	3597	...
QY	378	TyrAspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSer	397	...
DB	3598	ACTGATGAG	...	ATTACAGGAAGCCAGTTTGATCTACT 3636
QY	398	AsnIleLysTrpIleAspAlaProAspCysPheHisLeuTrpAsnAlaTrpGlu	417	...
DB	3637	GGTTTCCCATAT	...	3648
QY	418	GluProGluThrAspGluValValIleGlySerCysMetThrProProAspSerIle	437	...
DB	3649	GATCCATCTACAGGTGAATATTCGATCTCGAACTAAATTAATTCAGATCCAGATCAGTT	3708	...
QY	438	PheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLys	457	...
DB	3709	GCAGGTGAT	...	GAATCTCTCACTGAAGTA---TTGAACATTACA 3747
QY	458	ThrGlyLysSerThrArgArgProIleIleSerAsnGluAspGlnGlnValAsnLeuGlu	477	...
DB	3748	ACAGATGAAGTAAACAGGTTTGCCTCAAT	...	GATCTTGAA 3783
QY	478	AlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeu	497	...



Db	3784	ACTGGTCTT-----	3792
QY	498	AlaGluProTyrProLysValSerGlyPheAlaLysVal-----	AspLeuThrThr 514
Db	3793	-----CCAAAGAGATCCAGATATCAGGACTCCACAACTTCCAAATGGTACCTTGGTTGAT	3846
QY	515	GlyGluValLysHisLeuTyrGlyAspAsnArgTyrGlyGluProLeuPheLeu	534
Db	3847	CCATCAATAAATAAACCAATTCCAGTTCA-----	CATTCCGGA-----TTTATT 3891
QY	535	ProGlyGluGlyGlyGluGlu	541
Db	3892	AATGTACATCTGGAGACAA	3912

Search completed: November 17, 2003, 01:27:10  
Job time : 8102 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 16, 2003, 18:32:41 ; Search time 431 seconds  
(without alignments)  
3751.655 Million cell updates/sec

Title: US-09-758-269-6  
Perfect score: 3150  
Sequence: 1 MASPTATAAAGSRMLGNHNT.....VPYGFHGTGAGDILAKQVV 599

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2.1/UGPTO spool/US09758269/runat\_14112003.192308.25751/app query.fasta\_1.775  
-DB=N Geneseq 19Jun03 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTPMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09758269 @CGN\_1.0 @runat\_14112003.192308.25751 -NCPU=6 -ICPU=3  
-NO MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 19Jun03.\*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*
- 25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3150	100.0	1800	22	AAD09396 Arabidopsis thalia
2	2280.5	72.4	1818	22	AAD09401 Lycopersicon escul
3	2168.5	68.8	1839	22	AAD09399 Vigna unguiculata
4	1991	63.2	1752	22	AAD09394 Arabidopsis thalia
5	1930	61.3	1815	22	AAD09400 Zea mays neoxanthi
6	1663.5	52.8	1734	22	AAD09398 Arabidopsis thalia
7	969	30.8	1788	21	AAC42989 Arabidopsis thalia
8	969	30.8	1788	22	AAD09395 Arabidopsis thalia
9	969	30.8	1788	24	AB213639 cDNA encoding sunf
10	960.5	30.5	1950	22	AAF77206 Arabidopsis thalia
11	939	29.8	1777	21	AAC36083 Arabidopsis thalia
12	938	29.8	1617	22	AAD09397 Arabidopsis thalia
13	937	29.7	1617	22	AAD09402 Arabidopsis thalia
14	608	19.3	443	21	AAC56678 Eucalyptus grandis
15	592	18.8	492	22	AAC82706 Rice abscisic acid
16	556	17.7	491	21	AAC57157 Pinus radiata tran
17	522	16.6	393	25	ABX20484 Human GDP-mannose
18	511	16.2	372	21	AAC56695 Eucalyptus grandis
19	494.5	15.7	4403765	22	AAI93683 Mycobacterium tube
20	494.5	15.7	4411529	22	AAI93682 Mycobacterium tube
21	475.5	15.1	4403765	22	AAI93683 Mycobacterium tube
22	475.5	15.1	4411529	22	AAI93682 Mycobacterium tube
23	463	14.7	325	21	AAC56548 Eucalyptus grandis
24	415	13.2	386	21	AAC57162 Pinus radiata tran
25	409.5	13.0	340	21	AAC56520 Eucalyptus grandis
26	401	12.7	329	21	AAC57165 Pinus radiata tran
27	398	12.6	326	21	AAC57167 Pinus radiata tran
28	371.5	11.8	398	21	AAC57145 Human GDP-mannose
29	369	11.7	320	25	ABX20813 Pinus radiata tran
30	337.5	10.7	238	21	AAC57169 Eucalyptus grandis
31	330.5	10.5	567	21	AAC56546 Arabidopsis thalia
32	322.5	10.2	1713	22	AAD19635 Apoptosis inhibito
33	322	10.2	470	24	AB157537 Pinus radiata tran
34	311	9.9	246	21	AAC57168 Pinus radiata tran
35	311	9.9	412	21	AAC57164 Arabidopsis thalia
36	307.5	9.8	1791	22	AAD19634 Pinus radiata tran
37	290	9.2	308	21	AAC57142 Pinus radiata tran
38	286.5	9.1	395	21	AAC57117 Physcomitrella pat
39	279	8.9	534	22	AAH44254 Zea mays DNA fragm
40	272.5	8.7	190	21	AAC57163 Eucalyptus grandis
41	272.5	8.7	501	21	AAC43547 Beta, Beta-caroten
42	266.5	8.5	508	21	AAAS3888 Nucleotide sequenc
43	260	8.3	3111	21	AAH42170 Eucalyptus grandis
44	256	8.1	1773	22	AAH42170 Eucalyptus grandis
45	252.5	8.0	447	21	AAC56480

ALIGNMENTS

RESULT 1

AAD09396

ID AAD09396 standard; cDNA; 1800 BP.

XX AAD09396;

AC AAD09396;

DT 10-SEP-2001 (first entry)

XX Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED3 cDNA.

KW Neoxanthin cleavage enzyme; AtNCED3; abscisic acid; ABA; herbicide;

KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;

XX plant growth protectant; ss.

XX Arabidopsis thaliana.

XX Key Location/Qualifiers

FX

CDS 1..1800  
 FT FT /\*tag= a  
 XX /product= "Arabidopsis thaliana AtNCED3 protein"  
 PN EP11616794-A2.  
 XX 18-JUL-2001.  
 PD 11-JAN-2001; 2001EP-0300218.  
 XX PR 13-JAN-2000; 2000JP-0010056.  
 PR 11-JAN-2001; 2001JP-0003476.  
 XX (RIKE ) RIKEN KK.  
 PA  
 XX Iuchi S, Kobayashi M, Shinozaki K;  
 PI WPI; 2001-400081/43.  
 DR P-PSDB; AAE04784.  
 XX  
 PT A DNA encoding a protein with a neoxanthin cleavage activity for  
 PT producing transgenic plants with improved or decreased stress tolerance  
 PT  
 XX Claim 3; Page 32-36; 101pp; English.  
 XX  
 CC The invention relates to neoxanthin cleavage enzymes and their  
 CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key  
 CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.  
 CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a  
 CC plant when expressed in a plant cell. The invention also relates to  
 CC methods for increasing or decreasing stress tolerance in a plant by  
 CC introducing the DNA into the plant, and a transgenic plant into which a  
 CC neoxanthin cleavage enzyme is introduced. The improvement of stress  
 CC tolerance in plants is useful, for example in plant breeding. Neoxanthin  
 CC cleavage enzyme genes are useful for producing transgenic plants. An arid  
 CC land can be improved by growing transformant weed for several years and  
 CC then removing the weed by specifically lowering stress tolerance in the  
 CC weed by inducing an inducible promoter. The present cDNA sequence encodes  
 CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED3 protein.  
 CC The AtNCED3 cDNA is obtained from an Arabidopsis plant-derived cDNA  
 CC library using a cDNA of the cP965 (Cowpea Responsive to Dehydration)  
 CC gene isolated from cowpea plant as a probe.  
 XX  
 SQ Sequence 1800 BP; 458 A; 464 C; 439 G; 439 T; 0 other;

Alignment Scores:  
 Pred. No.: 3,92e-298 Length: 1800  
 Score: 3150.00 Matches: 599  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 22 Gaps: 0

US-09-758-269-6 (1-599) x AAD09396 (1-1800)

Qy 1 MetAlaSerPheThrAlaThrAlaAlaValSerClyVArgTrpLeuGlyGlyAenHisThr 20  
 Db 1 ATGGCTTCTTTACGGCAACGGCTGCGGTTCTGGGAGATGGCTTGTGGCAATCACTACT 60  
 Qy 21 GlnProProLeuSerSerSerGlnSerSerAspLeuSerTyrCysSerSerLeuProMet 40  
 Db 61 CAGCGGCATTATCGTCTCTCAAGCTCCGACTTGAGTTAGTTATGTAGCTCTTACCTATG 120  
 Qy 41 AlaSerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProAlaLeu 60  
 Db 121 GCCAGTCGTGTCACACGTAAGCTCAATGTTTCATCTCGGCTTCACACTCTCCAGCTCTT 180  
 Qy 61 HisPheProLysGlnSerSerAsnSerProAlaIleValValLysProLysAlaLysGlu 80  
 Db 181 CATTTCCCTAGCAATCATCAAACTCTCCCGCCATTGTTGTTAAGCCCAAGCCAAAGAA 240  
 Qy 81 SerAsnThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaLeuAspAlaAla 100

Db 241 TCCAAACACATAAACAGATGAATTTGTTCCAGAGAGCGCGCGGACGCTTGACGCGCGG 300  
 Qy 101 GluGlyPheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSer 120  
 Db 301 GAGGGTTTCTCTGTAGCCACGAGAAAGTACACCGCTTCTTAAAAACGGCTGATCTAGT 360  
 Qy 121 ValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgArgAsnLeuPro 140  
 Db 361 GTTCAGATCGCCGAAATTTGCTCCGGTGAATGAACACGCCCGCTCCGGGTATCTTCG 420  
 Qy 141 ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnPro 160  
 Db 421 GTGGTCGGAACCTTCCCGATTCCATCAAGAGAGTGTATGTGCGCAACGAGCTAACCCA 480  
 Qy 161 LeuHisGluProValThrGlyHisHisPhePheAspGlyAspGlyMetValHisAlaVal 180  
 Db 481 CTTACAGACCGGTGACAGTCCACCTTCTTCGACGAGACGATGGTATGGTTTCACCGGTC 540  
 Qy 181 LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal 200  
 Db 541 AAATTCGAACACGGTTTCAGCTAGCTAGCTTGCCTGCGGTTTACTCAGACTAACCGGTT 600  
 Qy 201 GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHis 220  
 Db 601 CAGGAACGTCAATTTGGGTCCGACCGGTTTCCCAAGACCATCGGTGAGCTTCACGGCCAC 660  
 Qy 221 ThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAspPro 240  
 Db 661 ACCGTTATTTCCCGACCTCATGCTATTCTACGCCAGAGCTGCACCGGTATAGTCGACCG 720  
 Qy 241 AlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAla 260  
 Db 721 GCACACGAAACCGGTGTAGCTAAACCGGTTTGGTCTATTTCATTCACCGCGGTATTGGCT 780  
 Qy 261 MetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThr 280  
 Db 781 ATGTCGAGAGGATGTTTACCTTACCAAGTTCAAGTCACTCCCAATGGAGATTTAAAAACC 840  
 Qy 281 ValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysVal 300  
 Db 841 GTTGGTGGTTTCGATTGTTGAGACAATTAGAATTCACAATGATGTCGCCACCGCGAAAGTC 900  
 Qy 301 AspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeu 320  
 Db 901 GACCCGGAATCCGGTGAACTCTTCGCTTAAAGCTACGACGCTGTTCCAAAGCTTACTTA 960  
 Qy 321 LysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAsp 340  
 Db 961 AAATACTTTCGATTCTCACCGACGGAATAAATCACCGGACGTCGAGATTCAGCTTGAT 1020  
 Qy 341 GlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValValProAspGln 360  
 Db 1021 CAGCCACGATGATGCACGATTTCCGATTTACAGAACTTCGTCGTCGACCTGACCCAG 1080  
 Qy 361 GlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyrAspLys 380  
 Db 1081 CAAAGTCGTTTTCAAGTCGCGAGATGATCCGCGTGGGTCTCCGCTGGTTTACGACAAG 1140  
 Qy 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnLys 400  
 Db 1141 AACAGGTGCGCAAGATTCGGATTTAGACAAATACCGCGAAGATTTCATCGAACTTAAG 1200  
 Qy 401 TrpIleAspAlaProAspCysPheHisLeuTrpAsnAlaIleTrpGluGluProGlu 420  
 Db 1201 TGGATTGATGCTCCAGATTGCTTCTGCTTCCATCTCTGGAACGCTTGGGAAGACCGAA 1260  
 Qy 421 ThrAspGluValValValIleGlySerCysMetThrProAspSerIlePheAsnGlu 440  
 Db 1261 ACAGATGAAGTCGTGATAGGTCTCTGTTATGCTCCACGAGACTCAATTTTCAACGAG 1320  
 Qy 441 SerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGlu 460

Db 1321 TCTCAGGAGATCTCAAGAGTGCTCTCTCAATCCGCTGAATCTCAAAACCGGTGAA 1380  
 Qy 461 SerThrArgArgProIleIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMet 480  
 Db 1381 TCRACTCGCGTCGATCATCTCCACGAAAGATCAACAAGTCAACTCGAAGCAGGATG 1440  
 Qy 481 ValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluPro 500  
 Db 1441 GTCACAGAAACATGCTCGCGCTAAACCAATTCGCTTACCTTGGCTTTAGCGAGCGG 1500  
 Qy 501 TrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHis 520  
 Db 1501 TGGCCTAAAGTCTCAGGATTCGCTAAAGTTGATCTCACTACTCGAGAAGTTAAAGAACAT 1560  
 Qy 521 LeuTyrGlyAspAsnArgTyrGlyGluProLeuPheLeuProGlyGluGlyGlu 540  
 Db 1561 CTTTACGGGATACCGGTTACGGAGAGAGCCCTCTGTTCTCCCGAGAGAGAGAG 1620  
 Qy 541 GluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLysSerGluLeu 560  
 Db 1621 GAAGCAGAGGATACATCTCTGTTTCGTTCAACGAGAGAGATCGAAATCGAGATTA 1680  
 Qy 561 GlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArgVal 580  
 Db 1681 CAGATAGTTAACGCGGTTAGCTTAGAGGTTGAAGCAACGGTTAACTTCGCTCAAGGTT 1740  
 Qy 581 ProTyrGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGlnValVal 599  
 Db 1741 CCGTACGAGATTACCGGTACATTCATCGAGCCGATGATTGGCAGACGAGTCTGTG 1797  
 RESULT 2  
 ID AAD09401 standard; cDNA; 1818 BP.  
 AC AAD09401;  
 XX  
 XX 10-SEP-2001 (first entry)  
 DT  
 XX Lycopersicon esculentum neoxanthin cleavage enzyme, LeNCED1 cDNA.  
 DE  
 XX Tomato; neoxanthin cleavage enzyme; LeNCED1; abscisic acid; ABA;  
 KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;  
 KW plant growth protectant; herbicide; ss.  
 XX  
 OS Lycopersicon esculentum.  
 XX  
 FH Key Location/Qualifiers  
 CDS 1..1818  
 FT /\*tag= a  
 FT /product= "Lycopersicon esculentum LeNCED1 protein"  
 XX  
 EP1116794-A2.  
 XX  
 PD 18-JUL-2001.  
 XX  
 PF 11-JAN-2001; 2001EP-0300218.  
 XX  
 PR 13-JAN-2000; 2000JP-0010056.  
 PR 11-JAN-2001; 2001JP-0003476.  
 XX  
 XX (RIKE ) RIKEN KK.  
 XX  
 XX Iuchi S, Kobayashi M, Shinozaki K;  
 XX WPI; 2001-400081/43.  
 DR P-PSDB; AB04789.  
 XX  
 XX A DNA encoding a protein with a neoxanthin cleavage activity for  
 PT producing transgenic plants with improved or decreased stress tolerance  
 PT  
 XX  
 PS Claim 3; Page 67-71; 101pp; English.  
 XX

CC The invention relates to neoxanthin cleavage enzymes and their  
 CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key  
 CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.  
 CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a  
 CC plant when expressed in a plant cell. The invention also relates to a  
 CC method for increasing or decreasing stress tolerance in a plant by  
 CC introducing the DNA into the plant, and a transgenic plant into which a  
 CC neoxanthin cleavage enzyme is introduced. The improvement of stress  
 CC tolerance in plants is useful, for example in plant breeding. Neoxanthin  
 CC cleavage enzyme genes are useful for producing transgenic plants. An arid  
 CC land can be improved by growing transformant weed for several years and  
 CC then removing the weed by specifically lowering stress tolerance in the  
 CC weed by inducing an inducible promoter. The present cDNA sequence encodes  
 CC lycopersicon esculentum neoxanthin cleavage enzyme, LeNCED1 protein  
 CC related to the invention.  
 XX  
 SQ Sequence 1818 BP; 569 A; 383 C; 369 G; 497 T; 0 other;

Alignment Scores:  
 Pred. No.: 5,178-213 Length: 1818  
 Score: 2280.50 Matches: 435  
 Percent Similarity: 82.04% Conservative: 63  
 Best Local Similarity: 71.66% Mismatches: 90  
 Query Match: 72.40% Indels: 19  
 DB: 22 Gaps: 6

US-09-758-269-6 (1-599) x AAD09401 (1-1818)  
 QY 5 ThrAlaThrAlaAlaValSerGlyArgTrpLeuGlyGlyAsnHisThrGlnProLeu 24  
 Db 7 ACTACTACTTCACATGCCCAAAATACATGATT-----AAGACTAAGTTG 51  
 QY 25 SerSerSerGlnSerSerAspLeuSerTyrCysSer-----SerLeuProMetAla 41  
 Db 52 TCAATGCCATCATCAAGGAGTTGGTTTTCATCAAACTCTATTCTCTACTCAAAAT 111  
 QY 42 SerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProAlaLeuHis 61  
 Db 112 CAACATAATAGCAAGTCTCAACATTAATCTCTCTCAAGCTCCACCTATATCTCAT 171  
 QY 62 PheProLysGlnSerSerAsn-----SerProAla-----IleValLysProLys 77  
 Db 172 TTTCCTAAACATCTCTCAAAATATCAACACCAAGAAATAATACAAATTTCAACCCAAA 231  
 QY 78 AlaLysGluSerAsn-----ThrLysGlnMetAsnLeuPheGlnArgAla 92  
 Db 232 CAAGAAACACAACTCTCTCTCTCACTTCCAGTGGATTTAGTCAGAAAGCA 291  
 QY 93 AlaAlaAlaLeuAspAlaAlaGluGlyPheLeuValSerHisGluLysLeuHisPro 112  
 Db 292 GCAGCAATGGCTTTAGATGCTGTAGAAAGTGCTTTAACTAAACATGAACCTTGAACACCC 351  
 QY 113 LeuProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAlaProValAsnGlu 132  
 Db 352 TTGCGAAACAGCCGACCCAGAGTCCAGATTTCTGGGAATTTTGCTCCGGTACCGGAA 411  
 QY 133 GlnProValArgAsnLeuProValValGlyLysLeuProAspSerIleLysGlyVal 152  
 Db 412 AATCCAGTCTGTCAATCTCTCCGGTCCCGGAAAAATACCCAAATGTGTTCAGGCGTT 471  
 QY 153 TyrValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHisPhePheAsp 172  
 Db 472 TAGGTTGAAACGGAGCTAACCCCTCTTTTGAACCAACCGCGGACACCAITTTCTGCAC 531  
 QY 173 GlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerTyrAlaCysArg 192  
 Db 532 GCGCAGGTTATGGTTACGCGCTTCAATTCAAAATGGTTCGGTAGTTACGTTGCCGT 591  
 QY 193 PheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPheProLys 212  
 Db 592 TTCACGTGAACACAGAGAGGCTTGTTCAGAAAAAGTTTGGGTGCGCCCTGTTTTCCTAAA 651  
 QY 213 AlaIleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyrAlaArg 232

652 GCATTTGGTGAATTAACATGCTCCTGGAATTCGAGGCTTATGCTTTTACGCTCGT 711  
233 AlaAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsnAlaGlyLeuVal 252  
712 GGGCTCTTCGGACTTGTGATCACAGTAAGAAAGTGGTGTGTCAAACGCCGGTTAGTC 771  
253 TyrPheAsnGlyArgLeuLeuAlaMetSerGluAspAspLeuProTyrGlnValGlnIle 272  
772 TATTTCAATCAACGATTAATCTGCTATGCTGAAAGATGATTTGCCTTACCATGTAAGGTA 831  
273 ThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeuGluSer 292  
832 ACACCCACCCGGCGATCTTAAACACAGAGGTCGATTCGATTTCCAGCGCCAGCTAAATCC 891  
293 ThrMetIleAlaHisProLysValAspProGluSerGlyGluLeuPheAlaLeuSerTyr 312  
892 ACCATGATAGCTACCCCAAGCTCGACCCAGTTTCGGTGACCTATTTCCTCTTAGCTAC 951  
313 AspValValSerLysProTyrLeuLysTyrPheArgPheSerProAspGlyThrLysSer 332  
952 GATGTGATTCAGAGCCATACCTCAAGTACTTCAGATTTTCAAAAAATGGGAAAAATCA 1011  
333 ProAspValGluIleGlnLeuAspGlnProThrMetMetHisAspPheAlaIleThrGlu 352  
1012 AATGATGTTGAATTCGATTCAGAGCCCAACCAATGATGATGATTCGCAATTAAGTAC 1071  
353 AsnPheValValProAspGlnGlnValValPheLysLeuProGluMetIleArgGly 372  
1072 AACTTCGTGCTCATCTTCATCAACAAAGTCGTTTTCAAGATGCTGMAATGATCCGTGA 1131  
373 GlySerProValValTyrAspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyr 392  
1132 GGTTCACCGTGTGTACGACAGAAAGATTTCCCGATTTGGTATTCTGGATAGATAC 1191  
393 AlaGluAspSerSerAsnIleLysTrpIleAspAlaProAspCysPheHisLeu 412  
1192 GCGAAGATGGTCTGATTTGAATGGTTGAAGTACCTGATTTGTTCTGTTCCACCTC 1251  
413 TrpAsnAlaTrpGluGluProGluThrAspGluValValIleGlySerCysMetThr 432  
1252 TGGAAATGCTTGGGAAGACAGAAACAGATGAATCGTTGTAATGCTTCATGTATGACA 1311  
433 ProProAspSerIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIle 452  
1312 CCACCACTCCATTTTCATGATGATGTGATGAGGCTTAAGAGTCTTTATCCGAAATC 1371  
453 ArgLeuAsnLeuLysThrGlyIleSerThrArgArgProIleLeuSerAsnGluAspGln 472  
1372 CGTCTCAATTTGAAACAGCGAAATCAACAGAAATCCATAATCGAAACCCCGGATGAA 1431  
473 GluValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPhe 492  
1432 CAAGTGAATTTAGAGCTGGATGGTGAACCGAAACAACTCGAAGGAAACAGAGTAT 1491  
493 AlaTyrLeuAlaLeuAlaGluProTyrProLysValSerGlyPheAlaLysValAspLeu 512  
1492 GCTTATTTGGTATCGCTGCAACATGCGCAAAAGTTTCTGGTTTGCAGAAAGTAAACCTG 1551  
513 ThrThrGlyGluValLysLysHisLeuTyrClyAspAsnArgTyrGlyGlyGluProLeu 532  
1552 TTCACCGGTGAAGTTGAGAAATTCATTTATGGTGACAAACAAATGATGGGGAACTCTT 1611  
533 PheLeuProGlyGlu-----GlyGlyGluGluAspGlyTyrIleLeuCysPheVal 550  
1612 TTTTACCAAGAGACCCCAACAGCAAGCAAGAGACGATGTTATATTTTAGCTTCGTT 1671  
551 HisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaValSerLeuGluVal 570  
1672 CACGATGAGAAGAAATGAAATCAGAACTGCAATTTGTAACGAAATGATGTTGAAGTTG 1731  
571 GluAlaThrValLysLeuProSerArgValProTyrGlyPheHisGlyThrPheLeuGly 590

1732 GAGGCAACTGTGAAGCTTCCATCAAGAGTTCTTATGATTTTCATGGAACATTCATAAAC 1791  
591 AlaAspAspLeuAlaLysGln 597  
1792 GCCATGATTTGGCAATCAG 1812

RESULT 3  
AAD09399  
ID AAD09399 standard; cDNA; 1839 BP.  
XX AAD09399;  
AC AAD09399;  
XX 10-SEP-2001 (first entry)  
XX Vigna unguiculata neoxanthin cleavage enzyme, CPRD65 cDNA.  
XX Cowpea; neoxanthin cleavage enzyme; abscisic acid; ABA; herbicide;  
XX stress tolerance; transgenic plant; plant breeding; antisense-therapy;  
XX plant growth protectant; CowPea Responsive to Dehydration; CPRD65; ss.  
XX Vigna unguiculata.  
OS  
XX Key Location/Qualifiers  
CDS 1..1839  
/\*tag= a  
/product= "Vigna unguiculata CPRD65 protein"

EP1116794-A2.  
18-JUL-2001.  
11-JAN-2001; 2001EP-0300218.  
13-JAN-2000; 2000JP-0010056.  
11-JAN-2001; 2001JP-0003476.  
(RIKE) RIKEN KK.  
Iuchi S, Kobayashi M, Shinozaki K;  
WPI: 2001-400081/43.  
P-FSDB; AAE04787.  
A DNA encoding a protein with a neoxanthin cleavage activity for  
producing transgenic plants with improved or decreased stress tolerance  
-  
Claim 3; Page 53-56; 101pp; English.  
The invention relates to neoxanthin cleavage enzymes and their  
corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key  
role in endogenous abscisic acid (ABA) biosynthesis under drought stress.  
Neoxanthin cleavage enzyme is used for improving stress tolerance in a  
plant when expressed in a plant cell. The invention also relates to  
methods for increasing or decreasing stress tolerance in a plant by  
introducing the DNA into the plant, and a transgenic plant into which a  
neoxanthin cleavage enzyme is introduced. The improvement of stress  
tolerance in plants is useful, for example in plant breeding. Neoxanthin  
cleavage enzyme genes are useful for producing transgenic plants. An arid  
land can be improved by growing transformant weed for several years and  
then removing the weed by specifically lowering stress tolerance in the  
weed by inducing an inducible promoter. The present cDNA sequence encodes  
Vigna unguiculata neoxanthin cleavage enzyme, CPRD65 (CowPea Responsive  
to Dehydration) protein. CPRD65 gene is isolated from cowpea plant.  
XX Sequence 1839 BP; 447 A; 595 C; 442 G; 355 T; 0 other;  
SQ

Alignment Scores:  
Pred. No.: 4,84e-202 Length: 1839  
Score: 2168.50 Matches: 423  
Percent Similarity: 78.44% Conservative: 61  
Best Local Similarity: 68.58% Mismatches: 98  
Query Match: 68.84% Indels: 35

DB: 22 Gaps: 8

US-09-758-269-6 (1-599) x AAD09399 (1-1839)

QY 8 AlaAlaValSerGlyArgTyrLeuGlyGlyAsnHisThrGlnProProLeuSerSerSer 27  
 : : : : :  
 Db 7 TCATCAGCTTCAAAACACTTGGTTTAACGCCCACTCCCATCTCCGCCCTTCAAAAGACCTA 66  
 : : : : :  
 QY 28 GlnSerSerAspLeuSerTyrCysSerSerLeuProMetAlaSerArgValThrArgLys 47  
 : : : : :  
 Db 67 CTTTCCACATCT---TCTCCCAAACTTACTTCTTTA-----AGAAAAACATCTCTCT 117  
 : : : : :  
 QY 48 LeuAsn---ValSerSerAlaLeuHisThrProProAlaLeuHisPheProLysGln--- 65  
 : : : : :  
 Db 118 TCCAAACACCATCACAATGTCCTTCAAAACA-----CTCCACTTCCCAAAACAGTAC 168  
 : : : : :  
 QY 66 -----SerSerAsnSerProAlaIleValLysProLysAlaLysGlnSer 81  
 : : : : :  
 Db 169 CAACCAACATCCACATCCATCCACAGCCACACACACACACCCCACTCAAACT 228  
 : : : : :  
 QY 82 AsnThr----- 83  
 : : : : :  
 Db 229 ACCACCATCACACACACACCGCCAGGAAACCAACCTCTCTCTGACACCAACAA 288  
 : : : : :  
 QY 84 -----LysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaAlaAlaAla 100  
 : : : : :  
 Db 289 CCATTACCTCAAAATGGAACTTCTCCAGAAAGCGCTGCCACCGCTTGGACCTGCTC 348  
 : : : : :  
 QY 101 GluGlyPheLeuValSerHisGlyLysLeuHisProLeuProLysThrAlaAspProSer 120  
 : : : : :  
 Db 349 GAAACCGCGCTGCTCGCACGAGCGCAACACCGCTCCCAAAACGGCGACCGAGG 408  
 : : : : :  
 QY 121 ValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgArgAsnLeuPro 140  
 : : : : :  
 Db 409 GTCCAAATCGCGGGAACCTTCGCGCGGTGCCGAGCATGCCCGGATCAAGACTCCCG 468  
 : : : : :  
 QY 141 ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnPro 160  
 : : : : :  
 Db 469 GTGGTGGAAAAATCCCAATGCTATGCGGGTGTACGTGGCBAACGGTCCCAATCG 528  
 : : : : :  
 QY 161 LeuHisGluProValThrGlyHisHisPheAspGlyAspGlyMetValHisAlaVal 180  
 : : : : :  
 Db 529 CTCTACGAGCCTGTGGCGCGGACCACTTCTTCGACGGCGAGCGCATGCTCCACCGCGTG 588  
 : : : : :  
 QY 181 LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal 200  
 : : : : :  
 Db 589 AAGTTCAGAACCGCGCGCGGACCTAGCGCTGCGCTGCTCCAGAGACGAGCGTCTCTCG 648  
 : : : : :  
 QY 201 GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHis 220  
 : : : : :  
 Db 649 CAGGAGAAATCTCTAGCGCGCGCGGTGTTCCCGAAGCCATCGGGAGCTCCACGCGCCAC 708  
 : : : : :  
 QY 221 ThrGlyIleAlaArgMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAspPro 240  
 : : : : :  
 Db 709 TCCGGCATCGCGCGGTCTCTCTCTACCGCGCGGTCTCTCTCGGGCTCGTTGATGGG 768  
 : : : : :  
 QY 241 AlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAla 260  
 : : : : :  
 Db 769 TCCAGGCGCATGGCGTGGCGAAGCGCGGTCTCGTCTACTTCAACACACCTCTTGGCC 828  
 : : : : :  
 QY 261 MetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThr 280  
 : : : : :  
 Db 829 ATGTCCGAACAGCATTTTACCTTACCCAGTGAGATCAACCTTAAACCGGACTTAAACACC 888  
 : : : : :  
 QY 281 ValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysVal 300  
 : : : : :  
 Db 889 GTTGGCGGTACGCTTCAACGGGACCTCAACTCAACATGATCGCCACCGCAACCTG 948  
 : : : : :  
 QY 301 AspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeu 320  
 : : : : :  
 Db 949 GACCCCGTCGACGGCGACCTCCACGCGTCAGCTACGACGTCACTCAGAAAGCTTACCTC 1008  
 : : : : :  
 QY 321 LysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAsp 340

Db 1009 AAGTACTTCGTTTCTCCCGGACGCGTCAAGTCCCGGAGTGAATCCCTGAG 1068  
 : : : : :  
 QY 341 GlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProaspGln 360  
 : : : : :  
 Db 1069 GACCCACCATGATGACCATTTTCCCAATACGAGAAATTTGTCGTCGTCGCGACCAAG 1128  
 : : : : :  
 QY 361 GlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyrAspLys 380  
 : : : : :  
 Db 1129 CAGGTGCTTTCAAACTAAGCGAGATGATCACCGCGGTCTCCCGTGTCTACGCAAG 1188  
 : : : : :  
 QY 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLys 400  
 : : : : :  
 Db 1189 AACAAACCTTCACGGTTTGGATTTTCACAGAATGCGAAGACGCGAATGCGATCGG 1248  
 : : : : :  
 QY 401 TrpIleAspAlaProaspCysPheCysPheHisLeuTrpAsnAlaTrpGluGluProGlu 420  
 : : : : :  
 Db 1249 TGGATCGACGCGCGGATTTGCTCTCCACTCTGGAACGCGTGGGAGGAGCCCGAA 1308  
 : : : : :  
 QY 421 ThrAspGluValValIleGlySerCysMetThrProProAspSerIlePheAsnGlu 440  
 : : : : :  
 Db 1309 ACCGAGAGGTTTGTGTGATTTGGTCTCTGATGACCCCTCGGACTCCATTTCAACGAA 1368  
 : : : : :  
 QY 441 SerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGlu 460  
 : : : : :  
 Db 1369 TCGGAGGAGATTTGAAGACGCTGCTCAGAGATAAGGCTGAATTTGAGACCGGCAAG 1428  
 : : : : :  
 QY 461 SerThrArgArgProIleIleSerAsnGluAspGlnValAsnLeuGluAlaGlyMet 480  
 : : : : :  
 Db 1429 TCCACTCGGCGCCCATTTATCTCC---GACGCGAACAAGTGAACCTGGGAAGCGGCGATG 1485  
 : : : : :  
 QY 481 ValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluPro 500  
 : : : : :  
 Db 1486 GTGNACAGAAACAAGCTCGGAGAGACCCAGTTCGCTATCTGGCTCTGGCGAGGCC 1545  
 : : : : :  
 QY 501 TrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHis 520  
 : : : : :  
 Db 1546 TGGCCCAAGTCTCGGCTTTGCAAAATGTTGTTGCTGAGTGGGAGTGAAGAAGTAC 1605  
 : : : : :  
 QY 521 LeuTyrGlyAspAsnArgTyrGlyGluProLeuPheLeuProGlyGlyGlyGlu 540  
 : : : : :  
 Db 1606 ATGTATCGAGAGAGAGTTCGTTGGGAGCCCTCTGTTCTTCCC---AACGCCCAAAA 1662  
 : : : : :  
 QY 541 GluAspGluGlyTyrIleLeuCysPheValHisAspGlyLysThrTrpLysSerGluLeu 560  
 : : : : :  
 Db 1663 GAACAGCATGGTATATTCTGGCATTCGTGCACGACGAGAAAGATGGAATCCGAGCTG 1722  
 : : : : :  
 QY 561 GlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArgVal 580  
 : : : : :  
 Db 1723 CAGATTGTGAATGCCAAAATTTAAGCTCGAACCTTCCATCAAACTCCCTCTCTGTT 1782  
 : : : : :  
 QY 581 ProTyrGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGln 597  
 : : : : :  
 Db 1783 CCTTACGTTTTCATGAACTTTCATTCATTCAGAGATTTGAGGAAACAA 1833  
 : : : : :

## RESULT 4

AAD09394  
 ID AAD09394 standard; cDNA; 1752 BP.

XX AC AAD09394;

XX DT 10-SEP-2001 (first entry)

XX DE Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCE1 cDNA.

KW Neoxanthin cleavage enzyme; AtNCE1; abscisic acid; ABA; herbicide;  
 KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;  
 KW plant growth protectant; ss.

OS Arabidopsis thaliana.

FX Location/Qualifiers  
 FT 1..1752







Db 694 GCAGCGCCGCGTGGCGCTGCTGGACCCCTCGCGCGGACCGCGGTGGCCAAACCGCGGC 753  
Qy 251 LeuValTyrPheAsnGlyArgLeuLeuAlaMetSerGluAspLeuProTyrGlnVal 270  
Db 754 CTGCTCTACTTCAACGCGCGCTGCTGCGCATGTCGAGGACGACCTCCCTTACCACGTC 813  
Qy 271 GlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeu 290  
Db 814 CGCGTGGCGACGACGCGGACCTCGAGACGCTGCGCGCTAGCACTTCAGCGGACGCTC 873  
Qy 291 GluSerThrMetIleAlaHisProLysValAspProGluSerGlyLeuLeuPheAlaLeu 310  
Db 874 GGCTGGCCCATGATCGGCACCCCAAGCTGGACCGCCGACCGGAGGAGCTCCACGCGTC 933  
Qy 311 SerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerProAspGlyThr 330  
Db 934 AGCTACGACGTCAAGAGCGCGTACCTCAAGTACTTCTACTTCAGCGCCGACGCGAC 993  
Qy 331 LysSerProAspValGluIleGlnLeuAspGlnProThrMetHisAspPheAlaIle 350  
Db 994 AAGTCCGACGCGTGGAGATCCGCTGGACGACCCACGATGATCCACGACTTGGCATC 1053  
Qy 351 ThrGluAsnPheValValProAspGlnGlnValPheLysLeuProGluMetIle 370  
Db 1054 ACCGAGAATTCGTGTGTGTCGCGACCCACCGGTGGTTCAGCTCCAGGAGATGCTG 1113  
Qy 371 ArgGlyGlySerProValValTyrAspLysAsnLysValAlaAArgPheGlyIleLeuAsp 390  
Db 1114 CGCGCGCGGTGCGCGCTGGTCTGACAGAGACGCTGCGGTTCGCGGTGCTCCCC 1173  
Qy 391 LysTyrAlaGluAspSerSerAsnIleLysTrpIleAspAlaProAspCysPheCysPhe 410  
Db 1174 AAGCACGCGCGGACGCGTGGAGATGGCGTGGGTGGACGCTGCGGACTGCTTCTGCTTC 1233  
Qy 411 HisLeuTrpAsnAlaTyrGluProGluThrAspGluValValIleGlySerCys 430  
Db 1234 CACTGTGGAAACGCTGGAGGACGAGCGGCGAGGTGGTGTGTGATCGGCTCTGTC 1293  
Qy 431 MetThrProAspSerIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSer 450  
Db 1294 ATGACCCCGCGGACTCCATCTTCAACGAGTCCGACGAGCGCTGGAGAGCGTGTGACC 1353  
Qy 451 GluIleArgLeuAsnLeuLysThrGlyLysThrArgProIleIleSerAsnGlu 470  
Db 1354 GAGATCGCGCTGGAGACGCGGCGCGGCGGTCCAGCGCGCGGCTCTGCGCG---CCG 1410  
Qy 471 AspGlnGlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThr 490  
Db 1411 TCGCAGCAGGAGAACCTGGAGGTGGCATGTGTGAACCGCAACCTGCTGGGCGCGAGC 1470  
Qy 491 LysPheAlaTyrLeuAlaLeuAlaGluProThrProLysValSerGlyPheAlaLysVal 510  
Db 1471 CGGTACGCGTACCTCGCGGTGGGAGCGGTGGCCCAAGAGTGGGCTTCGCCAAGGAG 1530  
Qy 511 AspLeuThrThrGlyGluValLysLysHisLeuTyrGlyAspAsnArgTyrGlyGlyGlu 530  
Db 1531 GACCTGTCACGCGGCGGACCTCACCAAGTTCGAGTACGCGGAGGCGCGGTTCGGCGCGAG 1590  
Qy 531 ProLeuPheLeuProGlyGluGlyGlyGlu-----GluAspGluGlyTyr 545  
Db 1591 CCCTGCTCTGCTCCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTAC 1650  
Qy 546 IleLeuCysPheValHisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAla 565  
Db 1651 GTGCTACCTTCGTCACGACGAGCGCGCGCGCGCGCGGTACCTGTTGTTCAATGCC 1710  
Qy 566 ValSerLeuGluValGluAlaThrValLysLeuProSerArgValProTyrGlyPheHis 585  
Db 1711 GCCGACATCCGCTGGAGCGACCGGTTCAGTGGCGTCCCGCGTCCGCTTCGCTTCCAC 1770  
Qy 586 GlyThrPheIleGlyAlaAspLeuAlaLysGln 597  
Db 1771 GGCACCTTCATCAGCGGCGCAGGAGCTCGAGGCCGAG 1806

RESULT 6  
AAD09398 standard; cDNA; 1734 BP.  
XX AAD09398;  
AC AAD09398;  
XX 10-SEP-2001 (first entry)  
DE Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED5 cDNA.  
XX Neoxanthin cleavage enzyme; AtNCED5; abscisic acid; ABA; herbicide;  
KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;  
KW plant growth protectant; ss.  
XX Arabidopsis thaliana.  
OS Arabidopsis thaliana.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..1734  
FT /\*tag= a  
FT /product= "Arabidopsis thaliana AtNCED5 protein"  
XX  
EP1116794-A2.  
XX 18-JUL-2001.  
XX 11-JAN-2001; 2001EP-0300218.  
XX 13-JAN-2000; 2000JP-0010056.  
PR 11-JAN-2001; 2001JP-0003476.  
XX (RIKE ) RIKEN KK.  
XX Tuchi S, Kobayashi M, Shinozaki K;  
XX WPI; 2001-400081/43.  
DR P-PSDB; AAE04786.  
XX  
PT A DNA encoding a protein with a neoxanthin cleavage activity for  
PT producing transgenic plants with improved or decreased stress tolerance  
PT  
XX  
Claim 3; Page 46-49; 101pp; English.  
XX  
PS The invention relates to neoxanthin cleavage enzymes and their  
CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key  
CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.  
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a  
CC plant when expressed in a plant cell. The invention also relates to  
CC methods for increasing or decreasing stress tolerance in a plant by  
CC introducing the DNA into the plant, and a transgenic plant into which a  
CC neoxanthin cleavage enzyme is introduced. The improvement of stress  
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin  
CC cleavage enzyme genes are useful for producing transgenic plants. An arid  
CC land can be improved by growing transformant weed for several years and  
CC then removing the weed by specifically lowering stress tolerance in the  
CC weed by inducing an inducible promoter. The present cDNA sequence encodes  
CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED5 protein.  
CC The AtNCED5 cDNA is obtained from an Arabidopsis plant-derived cDNA  
CC library using a cDNA of the CPD65 (CowPea Responsive to Dehydration)  
CC gene isolated from cowpea plant as a probe.  
XX  
SQ Sequence 1734 BP; 464 A; 390 C; 441 G; 439 T; 0 other;

Alignment Scores:  
Pred. No.: 1,2e-152 Length: 1734  
Score: 1663.50 Matches: 323  
Percent Similarity: 72.56% Conservative: 100  
Best Local Similarity: 55.40% Mismatches: 137  
Query Match: 52.81% Indels: 23  
DB: 22 Gaps: 9

US-09-758-269-6 (1-599) x AAD09398 (1-1734)

QY	24	LeuSerSerSerGlnSerSerAspLeuSerTyrCysSerSerLeuProMetAlaSerArg	43
DB	25	CTTCTTCGCGAAGACTTCTCTCGTTCTCATTTCTCCACAACCCAAAAATGCAAA	84
QY	44	ValThrArgLysLeuAsnValSerSer--AlaLeuHisThrProProAlaLeuHisPhe	62
DB	85	ATTTCTCGCGAATCTCATTAACCTTTCAAGATACCGACACTTCTCGTATCTCACTCT	144
QY	63	ProLysGlnSerSerAsnSerProAlaIleValValLysProLysAlaLysGluSerAsn	82
DB	145	CCGGTTCGGTCCGCG-----GTTAAGCTCAACCAAGTATCCAAAC	186
QY	83	ThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAlaIleGlu---	101
DB	187	-----TTAAACCTTTTCAAGACTTACGGCTACGATGCTCGCAAGATTGAGTCC	237
QY	102	GlyPheLeuValGerHisGluLysLeuHisProLeuProLysThrAlaAspProSerVal	121
DB	238	TCTATCTGTTATTCATATGGAGCAGATCGCCGCTTCTTAAACCGACCGACCCGGCGGT	297
QY	122	GluIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgArgAsnLeuProVal	141
DB	298	CAATATTCAGGTAACTTCGCTCGCGTTAAAGATGTCGCTTCAAGCGGTTAGAAAGT	357
QY	142	ValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnProLeu	161
DB	358	GTGGTTCAGATTCTCTTCTGTCFAAAGGAGTTTACATCCGTAACGGTCAACCCCTATG	417
QY	162	HisGluProValThrGlyHisHisPhePheAspGlyAspGlyMetValHisAlaValLys	181
DB	418	TTTCCGCGGTTAGCCGACATCATTTATTTACGCGTACGGAATGATTACGCGGTTAGT	477
QY	182	-----PheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPhe	199
DB	478	ATCGGTTTTCATAAC---CAGGTAGTTACAGCTCCGGTACACTAAAAAACAACCGGCTT	534
QY	200	ValGlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGly	219
DB	535	GTTCAAGAAACCGCGCTGGACGATCGGTTTCCCTAAACCAATCGCGGAGCTTCACGC	594
QY	220	HisThrGlyIleAlaArgLeuMetLeuPheTyrAlaAlaAlaAlaGlyIleValAsp	239
DB	595	CATTCCGGTCTAGTCTGACTCGCTCTTTTACGGCTCGAGTGGGATCGGCTAGTGGAC	654
QY	240	ProAlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeu	259
DB	655	GGGACACGTGGCATGGCGTAGCTAAACCGCGTGTGTTTCTTTAACGGCAGGTATTATTA	714
QY	260	AlaMetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLys	279
DB	715	GCCATGTCAAGAATGATCTCTCTTACAAGTGAAGATCGACGTCGAAGGAGATCTTGAG	774
QY	280	ThrValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLys	299
DB	775	ACGATCGGCGGTTCCGATTTCGATGACCGAGATTGACTTTCAGTGTAGCGCATCTTAAG	834
QY	300	ValAspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyr	319
DB	835	GTGGACCGCACCAAGAGATCTCCATACACTGAGCTACAAGTGTGAAGAACCTCAT	894
QY	320	LeuLysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeu	339
DB	895	CTCAGGTATCTTAATTCACACACGTGGGNAAGACACGTCAGCTGGAGATCACGCTC	954
QY	340	AspGlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAsp	359
DB	955	CCTGAACCAACGATGATTCATGATTTCGCGATAACCGAGAATTTTGTTCGTTATACCGAT	1014
QY	360	GlnGlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyrAsp	379
DB	1015	CAGCAATGGTATTCAAATATTCGAATATGATTCCGGCGCGGTCACCCGTTATCTAGCTT	1074

Qy	380	LysAsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIle	399
Db	1075	AAAGAAAAATGCGGAGATTGGAGTTTTTGTCAAGCAGGATCTGACCCGGTCCGGATATA	1134
Qy	400	LysTrpIleAspAlaProAspCysPheCysPheHisLeuTrpAsnAlaTrpGluGluPro	419
Db	1135	AATTGGTTCATGTACCGGATTCTTCTGTTCCATCTATGGAATCGGTGGGAAGAG--	1191
Qy	420	GluThrAspGlu-----ValValValIleGlySerCysMetThrProProAsp	435
Db	1192	AGAACCGAAGAGGAGACCCAGTTCGTCTAATCGGTCATGTATGAGCCACCCGAC	1251
Qy	436	SerIlePheAsnGluSerAspGluAenLeuLysSerValLeuSerGluIleArgLeuAsn	455
Db	1252	ACGATCTTTAGTGAATCAGAGAACACACCCGGTTGAATTAAGTGAGATCCGGTTAAAC	1311
Qy	456	LeuLysThrGlyGluSerThrArgAProIleIleSerAsnGluAspGlnGlnValAsn	475
Db	1312	ATCGGTACAAAAGAAATCGAACCGCTAAGGTTATCGTAACCGGA-----GTCAAT	1359
Qy	476	LeuGluAlaGlyMetValAenArgAsnMetLeuGlyVarGlyThrLysPheAlaTyrLeu	495
Db	1360	TTAGACGGGTTCACATAAACCGTAGTTACGTGGCGCGGAAGCCAGTTCTGTTTACATA	1419
Qy	496	AlaLeuAlaGluProTrpProLysValSerGlyPheAlaLysValAspLeuThrThrGly	515
Db	1420	GCAATAGCCGATCCTTGGCCCAAATGCAGTGGCATTGCGAAGCTAGATATACAAAACGGC	1479
Qy	516	GluValLysLysHisLeuTyrGlyAspAsnArgTyrGlyGlyGluProLeuPheLeuPro	535
Db	1480	ACGGTTTCAGAGTTAATTACGACGACGAGCGCGTTCGGTGGGAACCGTCTTTGTACCG	1539
Qy	536	GlyGluGlyGlyGluGluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThr	555
Db	1540	GAGGCAAGGAGAGAGAAGCAAAAGCTTATGTAATGGGTTCGTGAGAGACGAAGAGAA	1599
Qy	556	TrpLysSerGluLeuGlnIleValAsnAlaValSerLeuGluValGluAlaThrValLys	575
Db	1600	GACGAGTCGGAGTTTGTGGTGGTCGACGCGACGGATATGAAGCAAGTCGCGCGGTCCG	1659
Qy	576	LeuProSerArgValProTyrGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAla	595
Db	1660	TTGCCGAGAGGGTACCTTATGTTTCCATGGAAAGCTTCGTGAGCGAGATCAGTTGAAG	1719
Qy	596	LysGlnVal 598	
Db	1720	GAACAAGTT 1728	
RESULT 7			
AAC42389			
XX	ID	AAC42389 standard; DNA; 1788 BP.	
AC	AC	AAC42389;	
DT	DT	17-OCT-2000 (first entry)	
XX	XX	Arabidopsis thaliana DNA fragment SEQ ID NO: 37611.	
XX	XX	Hybridisation assay; genetic mapping; gene expression control;	
KW	KW	protein identification; signal transduction pathway;	
KW	KW	metabolic pathway; promoter; termination sequence; ss.	
XX	XX	Arabidopsis thaliana.	
XX	XX	EP1033405-A2.	
PN	PN	06-SEP-2000.	
XX	XX	25-FEB-2000; 2000EP-0301439.	
PF	PF	25-FEB-1999; 99US-0121825.	
PR	PR	05-MAR-1999; 99US-0123180.	
PR	PR	09-MAR-1999; 99US-0123548.	

PR 23-MAR-1999; 99US-0135788.  
PR 25-MAR-1999; 99US-0136264.  
PR 29-MAR-1999; 99US-0136785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0132866.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 25-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 18-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140931.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147453.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154778.  
PR 22-SEP-1999; 99US-0155133.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155653.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.

PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 22-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Alignment Scores: 1,21e-84 Length: 1788  
 Pred. No.: 969.00 Matches: 214  
 Score: 55.31% Conservative: 114  
 Percent Similarity: 36.09% Mismatches: 221  
 Best Local Similarity: 30.76% Indels: 44  
 Query Match: 21 Gaps: 13  
 DB: 21

US-09-758-269-6 (1-599) x AAC42989 (1-1788)

QY 32 LeuSerTyrCysSerSerLeuProMetAlaSerArgValThrArgLysLeuAsnValSer 51  
 DB 56 CTCTCTCCGCCCGCATCTTCCTCTC-----CTACTC 88  
 QY 52 SerAlaLeuHisThrProProAlaLeuHisPheProLysGlnSer-----SerAsn 68  
 DB 89 TCCTCCGTATCAACTCGCGCGTCGCGAAGACGTTCTCCAAATCACAAACCCCAAGGCACA 148  
 QY 69 SerProAlaIleValLysProLys-AlaLysGluSerAsnThrLysGlnMetAsnLe 88  
 DB 149 ACAATGATCGTGTAACAAACCCAAACACATCCACAAACCGAACCAATCAACCTTAGTCT 208  
 QY 88 uPheGlnArgAlaAlaAlaAlaAlaLeuAspAla-AlaGluGlyPheLeuValSerHisG 108  
 DB 209 CATCACCCGAACCTCCGACCAAGAAATGACTCTCGACACAGCTCTCTCACACCGCTCG 268  
 QY 108 LuLysLeu-----HisProLeuProLysThrAlaAspProSerValG 122  
 DB 269 AAGATGTAATCAACAGCTTCATCGATCCACCTTCACGCTCTCCGTTGATCCAAACATG 328  
 QY 122 LnlleAlaGlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuProValV 142  
 DB 329 TCCTCTGTATACTTCGCTTCCTGCTCTCGACAGCTTCTCCACAGAGCTGGAATCA 388  
 QY 142 al---GlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnProL 161  
 DB 389 TCACCGGCACCTCTCCACTGCTCACTTAACCGCGCTTACATCCGTAACGGTCCAAATCCAC 448  
 QY 161 euHisGluProValThrGlyHisPhePheAspGlyAspGlyMetValHisAlaValL 181  
 DB 449 AGTTCTCCCTCGTGTCTTACCATCTCTTCACGGCGACGGTATGCTTCACGCCATTA 508

QY 181 YSPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheValG 201  
 DB 509 AAATCCACAACCGTAAGCCACTCTCTGTAGCAGATACGTCAAGACTTATAATACACAG 568  
 QY 201 LngLuarGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHisT 221  
 DB 569 TCGAGAAACAAACCGGAGCTCCGTTATGCCTAACGTGTTTTCGGGATTCAACGGGTGTA 628  
 QY 221 hr---GlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAsp 240  
 DB 629 CGCGCTCAGTAGCTCGTGGAGCTTTAACCGCACCTAGGCTTTAACCGGACAGTATATC 688  
 QY 240 roAlaHisGlyThrGlyValAlaAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuA 260  
 DB 689 CGGTTAACGGCATTTGGTTAGCTTAACAGTCTAGCTTTCTTCAGTAACCGTCTCTTTG 748  
 QY 260 LametSerGluAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysT 280  
 DB 749 CTTTAGGTGAATCGATTATACCTACCGCTCGGTAACCGAATCAGGAGATATTGAAA 808  
 QY 280 hrValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysV 300  
 DB 809 CGATCGGACGTCAGATTTCGACGGGAATTAGCGATGAGTATGACAGCTCATCTTAAA 868  
 QY 300 alAspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrL 320  
 DB 869 CCGATCCATAAACCGGAGAACTTCGCTTTCGGGTACGTCGCGTT---CCACCGCTTT 925  
 QY 320 euLysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluIle---GlnL 339  
 DB 926 TAACATATTTCCGCTTTGATTCCGCCGGGAAACAAAGAGACGTTCCGATATTCGA 985  
 QY 339 euAspGlnProThrMetHisAspPheAlaIleThrGluAsnPheValValValProA 359  
 DB 986 TGACGTCTCCGCTCGTTCTCCATGACTTCGCGATCAGAAACGTCACGCGATTTCGAG 1045  
 QY 359 spGlnGlnValValPheLys-----LeuProGluMetIleArgGlyGlySerProV 376  
 DB 1046 AGATTCAGCTGGCATGAGGATGAACATGTTGGATTTCGGAAGGTGGTTCTCCGG 1105  
 QY 376 alValTyrAspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAsp 396  
 DB 1106 TTGTTACTGATAACCGGAAACCTCAAGGCTTGGAGTGATTCTTAAGTACGCCGAGATG 1165  
 QY 396 erSerAsnIleLysTyrIleAspAlaProAspCysPheCysPheHisLeuTrpAsnAla 416  
 DB 1166 AGTCGGAGATGAATGGTTCCGAAGTTCCTCGGATTCATCAATATCATTCACGCTATTAA 1225  
 QY 416 rpGluGluProGluThrAspGluValValIleGlySerCysMetThrProProAsp 436  
 DB 1226 GGGATGAAGATGATGGAACACGCGTCGTTTGTATGTCACCGAATATTATGTCGATTGA 1285  
 QY 436 erIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnL 456  
 DB 1286 ATACTTTAGAGAGATGGAT---CTGGTTTCATGTTTGGTGGAGAGGTGAAGATCGATC 1342  
 QY 456 euLysThrGlyGluSerThrArgArgProIleIleSerAsnGluAspGlnValAsnL 476  
 DB 1343 TCGTCACCGGATTTGAGAGCTCATCCGATCTCAGCGAGG-----AATC 1387  
 QY 476 euGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuA 496  
 DB 1388 TCGATTTCCGTGTGATTAAATCCCGCGGTTTCTCGGAGATGTAGCAGGTACGTTTACCGG 1447  
 QY 496 laLeuAlaGluProTrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyG 516  
 DB 1448 CGATTGGAGATCCGATCCCGAAGATCTCCGCTGGTGGTGAAGCTTGATGTCTTAAGAGAG 1507  
 QY 516 lu-----ValLysLysHisLeuTyrGlyAspAsnArgTyrGlyGlyGluP 531  
 DB 1508 ATCGGGATGATTGTACGCTGGCCCGTAGAATGTACGGTTACGGTTGTACGGCGGAGAAC 1567









PS Claim 1; Page 94-97; 135pp; English.

XX This invention relates to defence-related signalling genes isolated from the sunflower (*Helianthus annuus*). The genes encode a neoxanthin cleavage enzyme (NCE), an amino acid permease (AAP) and a glutamic acid rich protein (GRP). The signalling gene is useful for increasing the resistance of a plant to a pathogen such as fungus, virus, bacterium, nematode or insect (e.g. European corn borer), preferably *Sclerotinia* spp., *Phoma* spp., or *Phomopsis* spp., by stably incorporating a construct containing the gene into the genome of the plant. The gene is useful for regulating gene expression in a plant, in response to a stimulus such as infection with a pathogen, damage from a pathogen, hydrogen peroxide, jasmonic acid, methyl jasmonate, salicylic acid, oxalic acid or expression of a gene encoding oxalic acid oxidase. The genes are also useful for stem-preferred regulation of gene expression in a plant. The genes are useful in agriculture, particularly in the breeding of crop plants with improved agronomic traits, for modifying abscisic acid (ABA) metabolism and for modifying amino acid transport and content in plants. The present sequence represents cDNA encoding the sunflower neoxanthin cleavage enzyme (NCE).

XX Sequence 1950 BP; 535 A; 456 C; 453 G; 506 T; 0 other;

Alignment Scores:

Pred. No.:	9-37e-84	Length:	1950
Score:	960.50	Matches:	209
Percent Similarity:	53.9%	Conservative:	109
Best Local Similarity:	35.48%	Mismatches:	214
Query Match:	57	Indels:	57
DB:	22	Gaps:	13

US-09-758-269-6 (1-599) x AAF77206 (1-1950)

QY 51 SerSerAlaLeuHisThrProAlaLeuHisPheProLysGlnSerSerAsnSerPro 70  
 DB 49 TCACCTGTTCCACCCACACCA-----TCATCGTCTCACCG 87

QY 71 AlaileValVal-----LysProLysAlaLysGlnSer 81  
 DB 88 CCATCTCGCTAGAGTCTTTTCAGTCAGACTGAAGAAAAACACAGAACCGTCCACG 147

QY 82 -----AsnThrLysGln-----MetAsnLeuPheGln 90  
 DB 148 ACGCCACCCACCAAGAGACTAGTGACCGAAAGAACCAACCCCTGAATATTAGAAAA 207

QY 91 ArgAlaAlaAlaAlaLeuAspAla-----AlaGluGly 102  
 DB 208 CGTGAAGTTCAGTGTGGTGGATCAGTCTTTACCCCGCCACCATCTTTAATGCTTTGAT 267

QY 103 PheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSerValGln 122  
 DB 268 AGCATTAATAAATCTTTATGATCCACCGCGGAGGCTCTCAGTTGATCCAAAACACGTT 327

QY 123 IleAlaGlyAsnPheAlaProValAsnGluGlnProValArgAsnLeuProValVal 142  
 DB 328 TTCTCTGATAACTTTTCACCGGTGGAGCACTCCCTCCGACTGACTGTGAAGTCAATCGAG 387

QY 143 GlyLysLeuProAspSerLeuLysGlyValTyValArgAsnGlyAlaAsnProLeuHis 162  
 DB 388 GGCACACCTGCAAGTGTCTTGAAGTGTCTTACTTCGTAATGCTCGGAACCCGCAATTC 447

QY 163 GluProValThrGlyHisPhePheAspGlyAspGlyMetValHisAlaValLysPhe 182  
 DB 448 CTTCCCGAGGACCTTACCACTCTTCGATGGGATGGATGCTCATCTATTGATATC 507

QY 183 GluHisGlySerAlaSerTyAlaCysArgPheThrGlnThrAsnArgPheValGlnGlu 202  
 DB 508 TCCAATGGAAAGCTTCGTTATGATAGCCGATACATCAAAACATACAAATATTCAATAGAG 567

QY 203 ArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHisThr--- 221  
 DB 568 AAGAAGCGGGATGCCCATATTATCCAAACGCTTTTTCAGGGTTTAATGTGTGACTGCC---627

QY 222 GlyIleAlaArgLeuMetLeuPheTyAlaArgAlaAlaAlaGlyIleValAspProAla 241  
 DB 628 TCTGACGTCGATCGCAGTCACTCGCGCGCGATTTTGGCTGGACAATTTGACCCACCA 687

QY 242 HisGlyThrGlyValAlaAlaAlaGlyLeuValTyPheAsnGlyArgLeuLeuAlaMet 261  
 DB 688 AAGGTATTGGCTAGCCAAATACAGTCTGGCTTTTGGCAACAGACITTTTGTCTG 747

QY 262 SerGluAspAspLeuProTyGlnValGlnIleThrProAsnGlyAspLeuLysThrVal 281  
 DB 748 GGAGAGTCGGATCTCCCATATCGCTCAAACTAGCGCCGACGCTGACATAGTCAACGTC 807

QY 282 GlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysValAsp 301  
 DB 808 GGAGTGAAGACTTCGACGCAAACTATTTCATGAGCATGACCTCACCCAAAATTCAT 867

QY 302 ProGluSerGlyGluLeuPheAlaLeuSerTyAspValValSerLysProTyLeuLys 321  
 DB 868 CCAGTAACGAAGAAGCTTTTGGCTTTTGGTACGGTCCAGTC---CCCCCTTTCTTAACC 924

QY 322 TyrPheArgPheSerProAspGlyThrLysSerProAspValGluIle---GlnLeuAsp 340  
 DB 925 TTTTTCGGTTTCAACGAAAAACGAGAAAAACAAACGCGATGTCGCGATCTTCTCAATGACA 984

QY 341 GlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValValProAspGln 360  
 DB 985 AGCCCGTGGTCTTCACGACTTCGCAATCACCABAACTACCGGATTTTCCGAGATC 1044

QY 361 GlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyAspLys 380  
 DB 1045 CAAATCGGATGAGCCCAATGAGATGCTGGGTGGGGATCCCGGTTCAGCGGACGCT 1104

QY 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyAlaGluAspSerSerAsnIleLys 400  
 DB 1105 GGAAGGTGGCTCGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1164

QY 401 TrpIleAspAlaProAspCysPheCysPheHisLeuTrpAsnAlaTrpGluGluProGlu 420  
 DB 1165 TGGTTTGAAGTTCGGGTTTTTAATGTGATCATTCATCATCATCATCATCATCATCATCAT 1224

QY 421 ThrAspGluValValIleGlySerCysMetThrProProAspSerIlePheAsnGlu 440  
 DB 1225 GGAGATACGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1284

QY 441 SerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGlu 460  
 DB 1285 ATGGAT---TTGATTCATGATCGTGTGAGAAAGTGAAGATTAACTGTAACAGGAATG 1341

QY 461 SerThrArgArgProIleIleSerAsnGluAspGlnValAsnLeuGluAlaGlyMet 480  
 DB 1342 GTATCCCGCACCCGCTTTCACCCCGG-----AATCTTGATTTGGCGTT 1386

QY 481 ValAsnArgAsnMetLeuGlyArgGlyThrLysPheAlaTyLeuAlaLeuAlaGluPro 500  
 DB 1387 TTAATCCGCGCTTTGCTGCGTTTAAACACAGGTATATGTTGTGGAGTTGGTATCCG 1446

QY 501 TrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGlu----- 516  
 DB 1447 ATGCCAAAGATCTCTGTTGGTCAAGTAGATGTGCTACTCCGAGCAGACCGCTCGC 1506

QY 517 -----ValLysLysHisLeuTyGlyAspAsnArgTyGlyGlyGluProLeuPhe 533  
 DB 1507 GAATGCATAGTTGTCACCGGATGTTTGGCGCTTGTGTTACGTTGTTGAAACCATCTTT 1566

QY 534 LeuProGlyGlu-----GlyGlyGluGluAspGluGlyTyIleLeuCysPhe 549  
 DB 1567 GTAGCTAGGAGCGCAGCAACCCCGATGCGGATGAGGATGAGGATGAGGATGAGGATGAG 1626

QY 550 ValHisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaValSer----- 567  
 DB 1627 GTGATACAGAGAACACCGGTGAGTCAAGATTGCTGGTGTGATGAGCGCTAAGTACCCGAC 1686

QY 568 LeuGluValGluAlaThrValLysLeuProSerArgValProTyGlyPheHisGlyThr 587

```
Db      1687 CITGAGATCGTGCATCCGTGAAGCTGCCCCACCGGTACCATATGGTTTCATGGGTC 1746
Qy      588 PheilegLYAlaaspAspLeuAlaLys 596
Db      1747 TTGTGTAAGAGAAAGTGACGTAAACAG 1773

RESULT 11
AAC36083
ID AAC36083 standard; DNA; 1777 BP.
AC AAC36083;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 12487.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135623.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 08-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
```

PR	13-AUG-1999	99US-0148565	QY	64	LysGlnSerSerAsnSerProAlaIleValValValProLys---AlalysGluSerAsn	82
PR	13-AUG-1999	99US-0148684	QY	14	AACTCAGTGTGTCAGCAGCATCATCATCTCAGTCCATCTAGACCCCTCCAGGGTTTCTCC	73
PR	16-AUG-1999	99US-0149368	Db	83	ThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaAlaAlaAlaAlaAlaGluGly	102
PR	17-AUG-1999	99US-0149175	QY	74	TGGAAGCTTCTCGATCTTCTCGAGAGACTTGTGTCAAGCTCATG-----	118
PR	20-AUG-1999	99US-0149722	Db	103	PheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSerValGln	122
PR	20-AUG-1999	99US-0149929	QY	119	-----CAGCATGCTTCTCTCCCTCTCCAC-----TAC	145
PR	23-AUG-1999	99US-0149902	Db	123	IleAlaGlyAsnPheAlaProVal---AsnGluGlnProValArgAsnLeuProVal	141
PR	25-AUG-1999	99US-0150566	QY	146	CTCTCAGGCACTTCGCTCCCATCCGATGAACTCTCCCTCCGTCAGGATCTCCCGTC	205
PR	26-AUG-1999	99US-0150884	Db	142	ValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnProLeu	161
PR	27-AUG-1999	99US-0151065	QY	206	CATGGATTCTTCCGAATGCTTGAATGGTCAATTTGTAGGGTTGGTCCAAACCCCAAG	265
PR	27-AUG-1999	99US-0151080	Db	162	HisGluProValThrGlyHisHisPhePheAspGlyMetValHisAlaValLys	181
PR	30-AUG-1999	99US-0151303	QY	266	TTTGATGCTGTCGCTGGATATCACTGGTTGATGGAGATGGATGATTCATGGGTACGC	325
PR	31-AUG-1999	99US-0151438	Db	182	PheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheValGln	201
PR	01-SEP-1999	99US-0151930	QY	326	ATCAAAGATGGAAAGCTACTATGTTTCTCGATATGTTAAGACATCACGCTCTTAAGCAG	385
PR	07-SEP-1999	99US-0152363	Db	202	GluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHisThr	221
PR	10-SEP-1999	99US-0153070	QY	386	GAAGAAGTTCTCGAGCTGCCAAATTCATGNAG---ATTGGTACCTTAAGGGTTTTC	442
PR	13-SEP-1999	99US-0153758	Db	222	GlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAspProAla	241
PR	15-SEP-1999	99US-0154018	QY	443	GGATTGCTAATGCTCAATATCCAAACAGCTCAGAACGAGTCAAAATATTGGCAACACT	502
PR	16-SEP-1999	99US-0154039	Db	242	HisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuAlaMet	261
PR	20-SEP-1999	99US-0154779	QY	503	TATGGAATGGAACCTGCGCAATACAGCACTCGTATATCACCATGGAAACTTCTAGCATTA	562
PR	22-SEP-1999	99US-0155139	Db	262	SerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThrVal	281
PR	23-SEP-1999	99US-0155486	QY	563	CAGGAGGCAGATAAGCCGTACGTATCAAAAGTTTGGAGATGGAGACCTGCAAACTCTT	622
PR	24-SEP-1999	99US-0155659	Db	282	GlyArgPheAspPheAspGlyGlyLeuGluSerThrMetIleAlaHisProLysValAsp	301
PR	28-SEP-1999	99US-0156458	QY	623	GGTATAATAGATTATGCAAGAGATTGACCCACTCTTCTACTGCTGCCCAAAAGTTGAC	682
PR	29-SEP-1999	99US-0156596	Db	302	ProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeuLys	321
PR	04-OCT-1999	99US-0157117	QY	683	CCGGTTACGGGTGAAATGTTTACATTGCGCTATTTCG---CATACGCCACCTTATCTCACA	739
PR	05-OCT-1999	99US-0157753	Db	322	TyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAspGln	341
PR	06-OCT-1999	99US-0157865	QY	740	TACAGAGTTATCTCGAAAGATGGCATTTATGCATGCCAGTCCCAATCTATATCATCAG	799
PR	07-OCT-1999	99US-0158029	Db	342	ProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAspGlnGln	361
PR	08-OCT-1999	99US-0158232	QY	800	CCTATCATGATGATGATTGCTATTACTGAGACTTATGCAATCTTCTCATGGATCTTCTCT	859
PR	12-OCT-1999	99US-0158369	Db	362	ValValPheLysLeuProGluMetIleArgGlyGlySerProValVal---TyrAspLys	380
PR	13-OCT-1999	99US-0158393	QY	860	ATGCACCTTCAGGCCAAAGAAATGGTGAAGAGAGAAATATGATATATCTCATTTGATCCC	919
PR	13-OCT-1999	99US-0158394	Db	381	AsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLys	400
PR	13-OCT-1999	99US-0158295	QY	920	ACAAAAAGGCTCGTTTGGTGTCTTCCAGCTATGCCAGGATGAACCTTATGATTAGA	979
PR	14-OCT-1999	99US-0159329	Db	401	TrpIleAspAlaProAspCysPheCysPheHisIleuTrpAsnAlaTrpGluGluProGlu	420
PR	14-OCT-1999	99US-0159330	QY	980	TGTTTGGCTTCCCAACTGCTTATTTTCCACAAACGCAATGCTTGGAA-----GAA	1033
PR	14-OCT-1999	99US-0159637	Db			
PR	14-OCT-1999	99US-0159638	QY			
PR	18-OCT-1999	99US-0159584	Db			





```

US-09-758-269-6 (1-599) x AAD09402 (1-1617)
QY 64 LysGlnSerSerAsnSerProAlaIleValValysProLys---AlaLysGluSerAsn 82
DB 10 AAACCTCAGTATGCGCAGCATCATCTCAGTCCATCTAGACCCCTCCCAAGGTTCTCC 69
QY 83 ThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaAlaAlaAlaGluGly 102
DB 70 TCGAAGCTTCTCGATCTCTCGAGAGACTTGTCTGCAAGTCTCATG----- 114
QY 103 PheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSerValGln 122
DB 115 -----CACGATGCTTCTCTCCCTCTCCAC-----TAC 141
QY 123 IleAlaGlyAsnPheAlaProVal---AsnGluGlnProValArgAsnLeuProVal 141
DB 142 CTCTCAGGCAACTTCGCTCCCATCCCGTGATGAACTCCCTCCGTCAGGATCTCCCGGTC 201
QY 142 ValGlyLysLeuProAspSerIleLysGlyValIleValArgAsnGlyAlaAsnProLeu 161
DB 202 CATGGATTTCTTCCCGATGCTTCAATGGTGAATTTGTGAGGGTGGTCCCAACCCCAAG 261
QY 162 HisGluProValThrGlyHisPhePheAspGlyAspGlyMetValHisAlaValLys 181
DB 262 TTTGATGCTGTGCTGGATATCATCTGTTTATGAGATGGGATGATTCATGGGGTACGC 321
QY 182 PheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheValGln 201
DB 322 ATCAAGATGGGAAGTACTTATGTTCTCGATGTTTAAAGCATCATCAGCTCTTAAGCAG 381
QY 202 GluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHisThr 221
DB 382 GAAGAGTTCTTCGGAGCTGCGCAATTCATGAAG---ATTGGTGACCTTAAAGGGGTTTTC 438
QY 222 GlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAspProAla 241
DB 439 GGATTTGCTAATGGTCAATATCCACAGCTGAGAGAGAGCTCAAAATATTGGCAACACT 498
QY 242 HisGlyThrGlyValAlaAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAlaMet 261
DB 499 TATGGAATGGAAGTGGCAATACAGCAGCTCGTATATCACCATTGGAAGAACTTCTAGCATTA 558
QY 262 SerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThrVal 281
DB 559 CAGAGGCGAGATAAGCCGTGCTCATCAAGTTTGGAGATGGAGACCTGCAAACTCTT 618
QY 282 GlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysValAsp 301
DB 619 GGTATAATAGATTATGACAGAGATTGACCCACTCTTCACTGCTCACCACCAAAAGTTGAC 678
QY 302 ProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeuLys 321
DB 679 CCGGTATCGGGTGAATGTTTACATTCGGCTATTTCG---CATACGCCACCTTATCTCACA 735
QY 322 TyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAspGln 341
DB 736 TACAGAGTTATCTCGAAGATGGCATTATGATGACCCAGTCCCAATTACTATATACAGAG 795
QY 342 ProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAspGlnGln 361
DB 796 CCTATCATGATGATGATGATTTGCTTATTACTGAGACTTATGCAATCTTCATGAGTCTTCT 855
QY 362 ValValPheLeuProGluMetIleArgGlyGlySerProValVal---TyrAspLys 380
DB 856 ATGCATCTTCAGCCCAAGGAATGGTGAAGAGAGAAATGATATCTATCTATTGATCCC 915
QY 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerAsnIleLys 400
DB 916 ACAAAGGAGGCTCGTTTGGTGTCTTCGGGCTATGCCAAGGATGAACATTATGATTAGA 975
QY 401 TrpIleAspAlaProAspCysPheHisLeuTyrAsnAlaTyrGluGluProGlu 420

```

```

DB 976 TGGTTTGAGCTTCCCACTGCTTTATTTTCCACAAAGCCCAATGCTTGGGAA-----GAA 1029
QY 421 ThrAspGluValValIleGlySerCysMetThrProProAsp----- 435
DB 1030 GAGGATGAAGTCGCTCCCTCATCTGCTGCTTGGAGAAATCCAGATCTTGACATGGTCACT 1089
QY 436 SerIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsn 455
DB 1090 GGGAAAGTGAAGAAAGAACTCGAAATTTTGGCAACGAACCTGTACGAAATGAGATTCAAC 1149
QY 456 LeuLysThrGlyGluSerThrArgArgProIleIleSerAsnGluAspGlnGlnValAsn 475
DB 1150 ATGAAAACGGGCTCAGCTTCTCAAAAAAATACTCCCGCATCTCGC----- 1194
QY 476 LeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeu 495
DB 1195 GTTGATTTCCCAAGATCAATGAGTGTACACCGGAAGAAACAGAGATACGTATATGGA 1254
QY 496 AlaLeuAlaGluProTyrProLysValSerGlyPheAlaLysValAsp----- 511
DB 1255 ACAATTTCTGGACAGTATCGCAAGGTTTACCGGAATCATCAAGTTTGTATGTCATGCGAGAA 1314
QY 512 -----LeuThrThrGlyGluValLysLysHisLeuTyr----- 522
DB 1315 GCTGAGACAGGAAAGAAAGTCTCGAAGTAGGAGGTAAATATCAAGGAATATATGACCTG 1374
QY 523 GlyAspAsnArgTyrGlyGlyGluProLeuPheLeuProGlyGlyGlyGlyGluGluAsp 542
DB 1375 GGAGAAAGCAGATATGTTTCAGAGGCTATCTATCTCCGCTGAGACAGCAGAAAGAGAC 1434
QY 543 GluGlyTyrIleLeuCysPheValHisAspGluLysThrTyrLysSerGluLeuGlnIle 562
DB 1435 SAGCGTTTACTTGATATTTCTTTGTCATGATGAAACACACGGGAATCATGCGTCACTGTG 1494
QY 563 ValAsnAlaValSerLeuGluValGlu-----AlaThrValLysLeuProSerArgVal 580
DB 1495 ATAGACGCAAAACAAATCGCGCTCAACCGGTGAGTGTGGAGCTGCCGACAGGCTC 1554
QY 581 ProTyrGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGlnValVal 599
DB 1555 CCATATGCTTCCATGCTTGTGTTTGTACAGAGAAACAACTCCAGGAACAACTCTT 1611

RESULT 14
AAC56678
ID AAC56678 standard; DNA; 443 BP.
AC AAC56678;
XX
DT 25-JAN-2001 (first entry)
XX
DE Eucalyptus grandis transcription factor DNA sequence #549.
XX
KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW Poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW homeodomain zipper; LIM domain; AP2; ERBBs; zinc finger domain;
KW type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS Eucalyptus grandis.
XX
XX WO200053724-A2.
XX
XX 14-SEP-2000.
XX
XX 09-MAR-2000; 2000WO-US06112.
XX
XX 11-MAR-1999; 99US-0266513.
XX
XX 18-AUG-1999; 99US-0149485.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Wood M, McGrath A, Sherk MA, Glenn M;
PI

```



XX DR WPI; 2000-579369/54.  
 XX PT New isolated polynucleotide encoding a plant transcription factor for  
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
 PT having modified gene expression or modified activity of a polypeptide  
 XX  
 XX PS Claim 1; Page 490; 747pp; English.  
 XX  
 CC The present invention relates to novel plant transcription factors from  
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding  
 CC sequence for one such transcription factor. The transcription factor may  
 CC be used to produce a plant having modified gene expression such as a  
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or  
 CC mahogany species or to modify the activity of a polypeptide in a plant.  
 CC The transcription factors of the present invention are members from the  
 CC following families of regulatory proteins: bZIP, bZIP family of G-Box  
 CC binding factors, basic helix-loop-helix zipper,  
 CC homeotic/homeodomain/homeobox/WADS, homeodomain zipper, LIM domain, AP2  
 CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements  
 CC and MYB.  
 XX  
 SQ Sequence 443 BP; 71 A; 177 C; 117 G; 77 T; 1 other;  
 Alignment Scores:  
 Pred. No.: 3,38e-50 Length: 443  
 Score: 608.00 Matches: 115  
 Percent Similarity: 87.76% Conservative: 14  
 Best Local Similarity: 78.23% Mismatches: 18  
 Query Match: 19.30% Indels: 0  
 DB: 21 Gaps: 0  
 US-09-758-269-6 (1-599) x AAC56678 (1-443)  
 QY 208 ProValPheProLysAlaIleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMet 227  
 Db 3 CCGCTTCCNCAAGCCATCGCGAGCTCCACGCGCACTCCGCGCATCGCGGCTCATG 62  
 QY 228 LeuPheTyrAlaArgAlaAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAla 247  
 Db 63 CTTCTTACGCGCGGAGCTCTTCGCGCTCGCGACCCGCGAATGGCATGGCGTCCGG 122  
 QY 248 AsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAlaMetSerGluAspLeuPro 267  
 Db 123 AACGCGGCTGTGTACTTCGACGCCACCTCTTCGCGAATGTCGAGAGACGACTCCCC 182  
 QY 268 TyrGlnValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAsp 287  
 Db 183 TACCAGTCGGGTACGCGCTCCGCGACCTCGAGACCTCGAGCGCTCGCGCGCTACGACTTCGCC 242  
 QY 288 GlyGlnLeuGluSerThrMetIleAlaHisProLysValAspProGluSerGlyGluLeu 307  
 Db 243 GCGCAGCTCGATCTCCGATGATCCGCCACCCGAGATCGACCCCGGCTTCGCGGAGATG 302  
 QY 308 PheAlaLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerPro 327  
 Db 303 TTCGCCCTCAGCTACGACGTCGTCGGAAGCGGTACCTCAAGTACTTCGATTCCTCAAG 362  
 QY 328 AspGlyThrLysSerProAspValGluIleGlnLeuAspGlnProThrMetHisasp 347  
 Db 363 GACGGGGAAGATGCCCGGACGTCCGATCCCGCTCGAGATCCCGCTGGGTGACCCGACCATGACCAT 422  
 QY 348 PheAlaIleThrGluAsnPhe 354  
 Db 423 TTCGCCATCACCAGCGCTTT 443  
 RESULT 15  
 AAC82706/c  
 ID AAC82706 standard; DNA; 492 BP.  
 XX  
 AC AAC82706;  
 XX

DT 15-MAR-2001 (first entry)  
 XX  
 DE Rice abscisic acid synthesis associated DNA SEQ ID NO 5.  
 XX  
 XX Rice; abscisic acid synthesis; VP14-like gene; drought resistance; ds.  
 XX  
 OS Oryza sativa.  
 XX  
 PN WO200071727-A1.  
 XX  
 XX 30-NOV-2000.  
 PD  
 XX 25-MAY-1999; 99WO-JP02734.  
 PF  
 XX 25-MAY-1999; 99WO-JP02734.  
 PR  
 XX (NORU) JAPAN MIN AGRIC FORESTRY & FISHERIES  
 PA (BIOC-) BIO-ORIENTED TECHNOLOGY RES ADVANCEMENT.  
 PA  
 XX Hirochika H, Sakamoto K;  
 XX WPI; 2001-032042/04.  
 DR  
 XX  
 PT Oligonucleotide encoding gene for regulating abscisic acid synthesis in  
 PT plants, useful for constructing e.g. genetically-modified rice with  
 PT drought resistance and ear-germination resistance  
 XX  
 PS Disclosure; Page 50; 55pp; Japanese.  
 XX  
 CC This invention describes a novel polynucleotide sequence (I) which  
 CC encodes a protein capable of regulating the synthesis of abscisic acid.  
 CC The invention also describes (1) an oligonucleotide encoding a protein  
 CC of a gene obtained by controlling the expression of a VP14-like gene;  
 CC (2) a vector containing the oligonucleotide ligated operably to the  
 CC regulation sequence; (3) a plant transformed with the vector; and  
 CC (4) a method for regulating abscisic acid synthesis in a plant including  
 CC the transfer of the above oligonucleotide to it. The gene is useful for  
 CC constructing drought resistant rice.  
 XX  
 SQ Sequence 492 BP; 90 A; 126 C; 186 G; 88 T; 2 other;  
 Alignment Scores:  
 Pred. No.: 1,45e-48 Length: 492  
 Score: 592.00 Matches: 115  
 Percent Similarity: 90.28% Conservative: 15  
 Best Local Similarity: 79.86% Mismatches: 13  
 Query Match: 18.79% Indels: 2  
 DB: 22 Gaps: 0  
 US-09-758-269-6 (1-599) x AAC82706 (1-492)  
 QY 210 PheProLysAlaIleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPh 229  
 Db 434 TTTCTTAAGCGAATAGTGAGCTCCATGGCCACTCCGGGCAATCGCGGCTGCTGTGT 375  
 QY 229 eTyrAlaArgAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAla 249  
 Db 374 CTACGCGCGCGCGCTC-GGCTCTCTCGACCGCTCACACGCGCACCGGCTGCGCAACGC 316  
 QY 249 aGlyLeuValTyrPheAsnGlyArgLeuLeuAlaMetSerGluAspLeuProTyrG 269  
 Db 315 CGGCTCATCTACTTCAACGGCAGGCTCTCTCGCATGTCTGGAGGACGACTTCCCTTACCA 256  
 QY 269 nValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGly 289  
 Db 255 GGTGCGGCTCACCGCGCGGACCTCGAGACCGTCCGCGCTACGACTTCGACGGGCA 196  
 QY 289 nLeuGluSerThrMetIleAlaHisProLysValAspProGluSerGlyGluLeuPhe 309  
 Db 195 GCTCGGCTGCGCATGATCGCGCACCCCAAGTCGACCCCGGCGCACCGGAGCTCCACGC 136  
 QY 309 aLeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerProAsp 329

Mon Nov 17 10:00:59 2003

Db 135 GCTCAGCTACGACGTGATCAAGAAGCCGTACCTCAAGTACTTCTTCTCGCCCGACGG 76  
Qy 329 yThrLysSerProAspValGluIleGlnLeuAspGlnProThrMetMetHisAspPheAl 349  
Db 75 CACCAAGTCGGCGGACGTCGAGATCCCGCTCGACCCAGCCACCATGATCCAGACTTCGG 16  
Qy 349 alleThrGlu 352  
Db 15 AATTACTGAG 6

Search completed: November 16, 2003, 20:34:06  
Job time : 477 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 16, 2003, 20:09:17 ; Search time 6134 Seconds  
(without alignments)  
3994.924 Million cell updates/sec

Title: US-09-758-269-6  
Perfect score: 3150  
Sequence: 1 MASPTATAAAGSRWLGNNHT.....VPYGHGTFIGADDLAKQVV 599

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09758269/runat\_14112003\_192309\_25759/app\_query.fasta\_1.775  
-DB=GenEmbl -CPMT=fastap -SUFFIX=rg -MINMATCH=0.1 -LOOEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09758269 @CGN\_1\_1\_3963 @runat\_14112003\_192309\_25759 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOB=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

RESULT 1

ALIGNMENTS

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3150	100.0	1800	6	AX148306 Sequence
2	3150	100.0	1800	6	BD017431 Transgeni
3	3150	100.0	3869	8	AB026549 Arabidops
4	3150	100.0	52232	8	AB028617 Arabidops
5	3143	99.8	2331	8	AY056255 Arabidops
6	2291	72.7	2164	8	STU276244 Solanum t
7	2282.5	72.5	2171	8	LENEOXANT
8	2282.5	72.5	19018	8	LES439079
9	2280.5	72.4	1818	6	AX148316
10	2280.5	72.4	1818	6	BD017436
11	2172	69.0	2398	8	AF190462 Phaseolus
12	2170.5	68.9	95769	6	AX652128 Sequence
13	2170.5	68.9	95769	6	AC013430 Genomic s
14	2168.5	68.8	1839	6	AX148312
15	2168.5	68.8	1839	6	BD017434
16	2168.5	68.8	2432	8	AB030293 Vigna ung
17	2155.5	68.4	2357	8	AB080192 Pisum sat
18	2075.5	65.9	2034	8	AB080193 Pisum sat
19	2066.5	65.6	2310	8	AF224671 Persea am
20	2044.5	64.9	72058	8	AC074176 Arabidops
21	1991	63.2	1752	6	AX148302
22	1991	63.2	1752	6	BD017429
23	1991	63.2	110102	8	ATF28J12
24	1991	63.2	194143	8	ATCHRIV48
25	1988	63.1	175158	2	AC120531
26	1988	63.1	181159	2	CNS07YP7
27	1987	63.1	2153	8	AF224672
28	1962	62.3	133983	8	CNS08CB9
29	1959.5	62.2	131689	8	CNS08CB9
30	1930	61.3	1815	6	AX148314
31	1930	61.3	1815	6	BD017435
32	1930	61.3	2498	8	ZMU95953
33	1834.5	58.2	149097	2	AP005632
34	1663.5	52.8	1734	6	AX148310
35	1663.5	52.8	1734	6	BD017433
36	1663.5	52.8	78921	8	AB028621
37	1214	38.5	82697	8	ATT9A21
38	1014.5	32.2	120055	8	AC123596
39	1003	31.8	80686	2	AC144759
40	1003	31.8	120055	8	AC123596
41	986.5	31.3	2016	8	AB080194
42	971.5	30.8	80685	2	AC144759
43	969	30.8	1788	6	AX148304
44	969	30.8	1788	6	AX506749
45	969	30.8	1788	6	BD017430

AX148306  
LOCUS AX148306 1800 bp DNA linear PAT 08-JUN-2001  
DEFINITION Sequence 5 from Patent EP1116794.  
ACCESSION AX148306  
VERSION AX148306.1 GI:14347193  
KEYWORDS Arabidopsis thaliana (thale cress)  
SOURCE Arabidopsis thaliana  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
1  
REFERENCE  
1 Luchi, S., Kobayashi, M. and Shinozaki, K.  
Transgenic plants carrying neoxanthin cleavage enzyme gene  
Patent: EP 1116794-A 5 18-JUL-2001;  
Riken (JP)

## FEATURES

## source

1..1800 Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:3702"

## CDS

1..1800  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC41194.1"  
/db\_xref="GI:14347194"  
/translation="MASFTATAVSGRWLGNGHTQPLSSQSSDLSYCSSLPMSRV  
TRKLVNSALHTPPALHPKQSSNSPAIVVVKPAKESNTKOMNLFORAAALDAAG  
FLVSHKXLPKPADFSVQIAGNFAPVNEQVRNLFVVGKLPDSIKGVYRNGANP  
LHEVTHGHFDGGMVHAVKFEHGSAYACRTQTNRFVQERQLGRPPFPAIGSLH  
GHTGIASLMFLFYASAAAGIVDPAHGTGVANAGLVYFNGLRLAMSEDDLPYQGITENG  
DLKTVGRFPDQGLSEMIAPKVDPSGELFALSVDVSKPKLVKPRFSPDGTGKSPD  
VEIOLDPTWMDPFAITNFVVDPQVFKLPEMIRGSGPVVYDKNKKVARFGILDKY  
AEDSNKNIDAPCFCEHLNWAHEBETDEVVVGSCMTPPDSINESDENLKSVL  
EIRLNKGTSTRPILISNEDQVNLEAGVNRNMLGRKTKFAYLALAEWPVKVSGFA  
KVDLTTEVKKHLYGDRNYGEPLFPGEGEEDGEGYILCFVHDEKTKWSELQIVNAV  
SLVEATVKLPSRYPYGFHGTFFIGADLAKQVV"

BASE COUNT 458 a 464 c 439 g 439 t

## ORIGIN

## Alignment Scores:

Pred. No.: 6,73e-220 Length: 1800  
Score: 3150.00 Matches: 599  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-09-758-269-6 (1-599) x AX148306 (1-1800)

QY 1 MetAlaSerPheThrAlaThrAlaAlaValSerGlyArgTrrPLeuGlyGlyAsnHisThr 20  
DB 1 ATGGCTTCTTTTCCAGCGCAACGCTCGGCTTTCTGGAGATGGCTTGGTGGCAATCATACT 60  
QY 21 GlnProLeuSerSerGlnSerSerAspLeuSerTyrCysSerSerLeuProMet 40  
DB 61 CAGCGCCCAATTATCGCTTCTCTCAAGCTCCGACTTGAGTTATGAGCTCTCTTACCTATG 120  
QY 41 AlaSerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProProAlaLeu 60  
DB 121 GCCAGTCGTGTCCACAGTAAAGCTCAATGTTTCATCTGGCTTCACACTCCTCCAGCTCTT 180  
QY 61 HisPheProLysGlnSerSerAsnSerProAlaIleValValLysProLysAlaLysGlu 80  
DB 181 CATTTTCCCTAAGCAATCATCAAACTCTCCCGCATTTGTTTAAAGCCCAAGCCCAAGAA 240  
QY 81 SerAsnThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaAlaLeuAspAlaAla 100  
DB 241 TCCAACTAAACAGATGAATTTGTTTCCAGAGAGCGGCGCGCAGCTTGGACCGCGC 300  
QY 101 GluGlyPheLeuValSerHisGlnLysLeuHisProLeuProLysThrAlaAspProSer 120  
DB 301 GAGGTTTCTCTTGTCTAGCCAGCAGAGAGCTACACCGCTTCTTAAACCGCTGATCCTAGT 360

121 ValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgArgAsnLeuPro 140  
361 GTTCAGATCGCGGAAATTTTGTCTCCGCTGAATGAACAGCGCCGTCGCGCTTAATCTTCOG 420  
QY 141 ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyValAsnPro 160  
DB 421 GTGTGCGAAACATTCCTGATTCATCAAGAGATGTATGTGCGCAACGAGGTAAACCCA 480  
QY 161 LeuHisGluProValThrGlyHisPhePheAspGlyAspGlyMetValHisAlaVal 180  
DB 481 CTTCCAGAGCGGTGACAGAGTCACTTCTTCACGAGACGGTATGGTTCACGCGCTC 540  
QY 181 LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal 200  
DB 541 AATTCGAACACGGTTCAGCTAGCTAGCTTCCGCTTCCGCTTACTCAGACTTAACCGGTTGTT 600  
QY 201 GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHis 220  
DB 601 CAGGAACGTCAATTTGGGTGCGACCGGTTTCCCAAGCATCGGTGAGCTTCACGGCCAC 660  
QY 221 ThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaGlyIleValAspPro 240  
DB 661 ACCGGTATTGCCCGACTCATCTATTCTACGCCAGAGCTGCGCGGTATAGTTCGACCG 720  
QY 241 AlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuAla 260  
DB 721 GCACACGGAACCGGTGTAGCTAACCGCGTTTGTCTATTTCATGGCGGTATTGGCT 780  
QY 261 MetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThr 280  
DB 781 ATGTCCGAGGATGATTTTACCTTACCAAGTTCAGATCACTCCCAATGGAGATTTAAACCC 840  
QY 281 ValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysVal 300  
DB 841 GTTGTGCTGCTGATTTTGTATGGACAATTAGAAATCCACAATGATTGCCCAACCGAAGTC 900  
QY 301 AspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeu 320  
DB 901 GACCCGGAATCCGCGTGAATCTCTTCGTCTTAAAGTACGACGCTGTTTCAAGAGCTTACCTA 960  
QY 321 LysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAsp 340  
DB 961 AAATATCTCCGATTTCTCCCGACGGAATAATCACCGGAGCTCGAGATTCAGCTTGTAT 1020  
QY 341 GlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValProAspGln 360  
DB 1021 CAGCCAAACGATGATGCAGATTCGCGATTACAGAGAACTTCGCTCGTCTGCTACCTACCAG 1080  
QY 361 GlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyrAspLys 380  
DB 1081 CAGTCTGTTTTCAAGCTCCGAGATGATCCCGGTGGGTCTCCGCTGGTTCACGACAAG 1140  
QY 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnLysLys 400  
DB 1141 AACCAAGTCCGAAGATTCGGGATTTTAGACAAATACGCGAAGATTCATCGAACATTAAG 1200  
QY 401 TrpIleAspAlaProAspCysPheCysPheHisLeuTrpAsnAlaTrpGluGluProGlu 420  
DB 1201 TGGATTGATGCTCCAGATGCTTCTGCTTCCATCTCTGGAACGCTTGGGAAGCGCAGAA 1260  
QY 421 ThrAspGluValValIleGlySerCysMetThrProProAspSerIlePheAsnGlu 440  
DB 1261 ACAGATGAAGTCTGCTGATAGGCTCTGTATAGCTTCCACACGAGACTCAATTTTCAACGAG 1320  
QY 441 SerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGlu 460  
DB 1321 TCTGACGAGAAATCTCAAGAGATGCTCTGCTGAAATCCGCTGAAATCTCAAAACCGGTGAA 1380  
QY 461 SerThrArgArgProIleIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMet 480  
DB 1381 TCAACTCGCGCTCCGATCATCTCCAACGAAGATCAACAAGCTCAACCTCAAGCAGCGGATG 1440

```

QY 481 ValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluPro 500
DB 1441 GTCACAGAAAACAATGCTGGCGGTAAACCAATTCGCTTACTTGGCTTAGCCGAGCG 1500
QY 501 TrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHis 520
DB 1501 TGGCCCTAAAGTCTCAGGATTCGCTAAAGTTGATCTCACTACTGGAAGAGTTAAGAAACAT 1560
QY 521 LeuTyrGlyAspAsnArgTyrGlyGlyGluProLeuPheLeuProGlyGluGlyGlu 540
DB 1561 CTTTACGGCGATTAACCGTTACGGAGGAGAGCTCTGTCTCTCCCGGAGAGGAGAGAG 1620
QY 541 GluAspGluGlyTyrTrpLeuLysPheValHisAspGluLysThrTrpLysSerGluLeu 560
DB 1621 GAAGACCAAGGATACATCTCTGTTCGTTTCAGCAGAGACATCGAAATCGGAGTTA 1680
QY 561 GlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArgVal 580
DB 1681 CAGATAGTTTACGCCGTTTACGTTAGAGTTTGAAGCAACGTTTAACTTCCGTCAGAGTT 1740
QY 581 ProTyrGlyPheHisGlyThrPheIleGlyAlaAspLeuAlaLysGlnValVal 599
DB 1741 CCGTACGGATTCACGCTACATTCATCGAGCGGATGATTTGCGAAGCAGGTCGTG 1797

RESULT 2
BD017431
LOCUS Transgenic plant using neozanthine cleaving enzyme gene.
DEFINITION BD017431
ACCESSION BD017431
VERSION BD017431.1 GI:22558607
KEYWORDS JP 2001258579-A/3
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 1800)
AUTHORS Iuchi, K., Kobayashi, M. and Shinozaki, K.
TITLE Transgenic plant using neozanthine cleaving enzyme gene
JOURNAL Patent: JP 2001258579-A 3 25-SEP-2001;
COMMENT THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH
OS Arabidopsis thaliana (thale cress)
PN JP 2001258579-A/3
PD 25-SEP-2001
PF 11-JAN-2001 JP 2001003476
PI KIYOSHI IUCHI, MASATOMO KOBAYASHI, KAZUO SHINOZAKI PC
GI2N15/09,A01H5/00,C12N5/10,C12N9/02,C12N15/00,C12N5/00 CC
Transgenic plant using neozanthine cleaving enzyme gene FH
Location/Qualifiers
FT CDS Location/Qualifiers
source 1..1800
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
BASE COUNT 458 a 464 c 439 g 439 t
ORIGIN

Alignment Scores:
Pred. No.: 6 73e-220 Length: 1800
Score: 3150.00 Matches: 599
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-758-269-6 (1-599) x BD017431 (1-1800)
QY 1 MetAlaSerPheThrAlaThrAlaAlaValSerGlyArgTrpLeuGlyLysHisThr 20
DB 1 ATGGCTTCTTTCAGGCAACGCTGCGGTTCTTGGAGATGGCTTGGCGCAATCATACT 60
QY 21 GlnProLeuSerSerGlnSerGlnSerAspLeuSerTyrCysSerSerLeuProMet 40

```

```

DB 61 CAGCGCGCAATATCGCTCTTCTCAAGCTCCGACTTCAGTTATTGTAGCTCCTTACCTATG 120
QY 41 AlaSerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProProAlaLeu 60
DB 121 GCCAGTCTGTACACGTAAGCTCAATGTTTCACTGCGCTTCCACTCTCCAGCTCTT 180
QY 61 HisPheProLysGlnSerSerAsnSerProAlaIleValValLysProLysAlaLysGlu 80
DB 181 CATTTCCCTTAAGCAATCATCAAACTCTCCGCCCATTTGTTTAAAGCCAAAGCAAAGAA 240
QY 81 SerAsnThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaAlaLeuAspAlaAla 100
DB 241 TCCAACTAAACAGATGAATTTGTTCCAGAGAGCGCGCGCGAGCTTGGACGCGCGCG 300
QY 101 GlnGlyPheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSer 120
DB 301 GAGGTTTCTTGTGACGACGAGAGCTACACCGCTTCTCTAAACGGCTGATCTCTAGT 360
QY 121 ValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgArgAsnLeuPro 140
DB 361 GTTCAGATCCCGGAAATTTGCTCCGTTGAATGAACAGCCCGCTCCGGCGTAAATCTTCG 420
QY 141 ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyValaAsnPro 160
DB 421 GTGCTCGGAAACTTCCCGATTCCATCAAGAGAGTGTATGTGCGCAACGGAGCTAACCCA 480
QY 161 LeuHisGluProValThrGlyHisHisPhePheAspGlyAspGlyMetValHisAlaVal 180
DB 481 CTTTCAGAGCGGATGACAGGTCACCACTTCTTCGAGCGGACGATGATGTTTCACGCGTC 540
QY 181 LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal 200
DB 541 AAATTCGAACAGGTTTCAGTACTAGCTTTCGCGTTTACTCAGACTAACCGGTTTGT 600
QY 201 GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHis 220
DB 601 CAGGAACGTCATTTGGTGTGACCGGTTTCCCAAAAGCCATCGCTGAGCTTCACGCGCAC 660
QY 221 ThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAspPro 240
DB 661 ACCGTTATTCGCGGACTCATGCTATTCTACGCGAGAGCTGCACCGGTTATGTCGACCG 720
QY 241 AlaHisGlyThrGlyValAlaAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuAla 260
DB 721 GCACACGGAACCGGTGTAGCTAACCGCGTTTGGTCTATTTCATTCATGCGCGGTTATTGGCT 780
QY 261 MetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThr 280
DB 781 ATGTCGAGGATGATTTACCTTACCAAGTTTCAGATCACTCCCAATGGAGATTTAAAAACC 840
QY 281 ValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysVal 300
DB 841 GTTGGTTCGGTTTCGATTTTGTGACCAATTAGATCCACATGATGTCGCCACCGAAAGTC 900
QY 301 AspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeu 320
DB 901 GACCCGGAATCCGGTGAACCTTTCGCTTAAAGCTACGACGTCGTTTCAAAAGCTTACCTA 960
QY 321 LysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAsp 340
DB 961 AAATCTTCCGATTTCTCAGCGAGGAACTAAATCAGCGAGCTCGAGATTCAGTTGAT 1020
QY 341 GlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValValProAspGln 360
DB 1021 CAGCCAAAGATGATGACGATTTTCGATTTACAGAACTTCGTCGTCGTCGTCGTCGTCG 1080
QY 361 GlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyrAspLys 380
DB 1081 CAAGTCGTTTCAAGCTCCCGAGATGATCCCGCGTGGGTCCTCCGTTGGTTTACCAAG 1140
QY 381 AsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLys 400

```

```

Db      1141 AACAGGTCGCAAGATTCGGGATTTAGACAAATACGCCGAGAGATTCATCGAACATTAAAG 1200
Qy      401 TTPileAspAlaProAspCysPheCysPheHisLeuTTPAsnAlaTTPiGluProGlu 420
Db      1201 TGGATTGATGCTCCAGATTGCTTCTGCTTCATCTCTGGAACGCTTGGGAGAGCGAGAA 1260
Qy      421 ThrAspGluValValIleGlySerCysMetThrProProAspSerIlePheAsnGlu 440
Db      1261 ACAGATGAAGTCGCTGATAGGCTCCTGTATGACTCCACAGACTCAATTTTCAACGAG 1320
Qy      441 SerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGlu 460
Db      1321 TCTGACGAGAAATCTCAAGAGTGTCTGTCTGAATCCGCTGAATCTCAAAACCGGTGAA 1380
Qy      461 SerThrArgArgProIleLeuSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMet 480
Db      1381 TCAACTCGCCGTCGATCATCTCCACGAGAGATCAACAAGTCAACTCGAACAGGATG 1440
Qy      481 ValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluPro 500
Db      1441 GTCAACAGAAACATGCTCGCCGTAACAAACCAATTCGCTTACTTGGCTTTAGCCGAGCG 1500
Qy      501 TTPProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHis 520
Db      1501 TGGCCTAAAGTCTCAGGATTCTGCTAAAGTTGATCTCCTACTCGAGAAGTTAAGAAACAT 1560
Qy      521 LeuTyrGlyAspAsnArgTyrGlyGlyGluProLeuPheLeuProGlyGlyGlyGlu 540
Db      1561 CTTTACGCCGATTAACCGTTACGAGGAGAGACCTCTGTTCTCCCGCGAGAGAGAGAG 1620
Qy      541 GluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTTPlysSerGluLeu 560
Db      1621 GAAGACGAAGATACATCTCTGTTTCGTTTCACGACGAGAGACATGGAAATCGGAGTTA 1680
Qy      561 GlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArgVal 580
Db      1681 CAGATAGTTTAAACCGGTAGCTTAGAGTTGAGCAACGGTTAACTTCGTCGCAAGGTT 1740
Qy      581 ProTyrGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGlnValVal 599
Db      1741 CCGTACGGATTTCACGGTACATTCATCGGAGCCGATGATTTGGCGAAGCAGGTCGTG 1797

RESULT 3
AB026549 LOCUS
DEFINITION Arabidopsis thaliana gene for neoxanthin cleavage enzyme, complete cds.
ACCESSION AB026549
VERSION AB026549.1 GI:16416373
KEYWORDS neoxanthin cleavage enzyme.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE 1
AUTHORS Iuchi,S., Kobayashi,M. and Shinozaki,K.
TITLE Characterization of neoxanthin cleavage enzyme from Arabidopsis thaliana
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3869)
AUTHORS Iuchi,S. and Shinozaki,K.
TITLE Direct Submission
JOURNAL Submitted (22-APR-1999) Satoshi Iuchi, RIKEN, Plant Mol. Bio.; 3-1-1 Kouyadai, Tsukuba 305-0074, Japan
(E-mail:iuchi@rc.riken.go.jp, Tel:81-298-36-4359)
location/Qualifiers
FEATURES
source
1..3869
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Col."
/db_xref="taxon:3702"
1831..3630

```

```

/codon_start=1
/product="neoxanthin cleavage enzyme"
/protein_id="BAB70609.1"
/db_xref="GI:16416374"
/translation="MASPTATAAASVSGWLGNNHTPQLSSSSQSSDLSSYCSSLPMSARV
TRLNYSALHTPPALFPFKQSNSPAIVVPKAKESNTKQNLFORAAALDAABG
FLVSHKHLPLKPTADPSVQIAGNPAVNEQPVRRNLVYVKLPDPSIKGYVVRNAGP
LHPVTGHPFDGDMVHVKFPHGSAVACRFTQTRFVQERQLGRPVFPKAIAGLH
GHTGIALMLFYARAAGIVDPAHGTGVANAGLVYENGLLAMSEDLPYVOITPNG
DLKTVGSPFDGQLESTMTAHPKVDPESGELFALSVDVSVKPYLVKFRSPDGTKSPD
VEIQLODTHMDHFAITENFVVDPDQOVFKLPEMTRGSPVYVYDKNVARFGILDKY
AESSNLIKTDADPCFCFLWNAWPEDEIVVVGSCMTTPDSIINSESDENKSVLS
EILNLTGTESTRP.IINSEDOQNLEAGMVRNLGRKTFAYLALABFPWPKVSFA
KVLDTTGEVKHLYGDNRYGGEPLFLPGEGEDEGYILCFVHDEKTKWSELOVNAV
SLEVEATVKLPSPVPYGFHGTFGADLAKQVV"
BASE COUNT 1155 a 845 c 773 g 1096 t
ORIGIN
Alignment Scores:
Pred. No.: 1,99e-219 Length: 3869
Score: 3150.00 Matches: 599
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-09-758-269-6 (1-599) x AB026549 (1-3869)
Qy 1 MetAlaSerPheThrAlaThrAlaAlaValSerGlyArgTyrLeuGlyGlyAsnHisThr 20
Db 1831 ATGGCTTCTTTCACGCAACGGCTGCGGTTCTGGGAGATGGCTTGGTGGCAATCATACT 1890
Qy 21 GlnProProLeuSerSerSerGlnSerSerAspLeuSerTyrCysSerSerLeuProMet 40
Db 1891 CAGCGCGCATATTCGTTCTCTAAAGCTCCGACTTGAGTTATTGTAGCTCTTACCTATG 1950
Qy 41 AlaSerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProProAlaLeu 60
Db 1951 GCCAGTGTGTCACACGTAAGCTCAATGTTTCATCTCGCTTTCACACTCTCCAGCTCTT 2010
Qy 61 HisPheProLysGlnSerSerAsnSerProAlaIleValLysProLysAlaLysGlu 80
Db 2011 CATTTCCCTAAGCAATCATCAAACTCTCCGCGCAATTTGTTTAAAGCCCAAGCAAGAA 2070
Qy 81 SerAsnThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAlaAla 100
Db 2071 TCCACACATAACAGATGAATTTGTTCCAGAGACGGCGCGGCGGCGGCGGCGGCGG 2130
Qy 101 GluGlyPheLeuValSerHisGlyLysLeuHisProLeuProLysThrAlaAspProSer 120
Db 2131 GAGGGTTTCTTGTGTCAGCCAGAGAGCTACACCGCTTCTCTAAACCGGCTGATCTTAGT 2190
Qy 121 ValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgArgAsnLeuPro 140
Db 2191 GTTCAGATCGCGGGAATTTTGTCCGGTGAATGACAGCCCGTCCGCGGTAATCTTCCG 2250
Qy 141 ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyValAsnPro 160
Db 2251 GTGGTCGGAACACTTCCCGATTCCATCAAGAGAGTGTATGTGCGCAACGGAGCTAACCA 2310
Qy 161 LeuHisGluProValThrGlyHisHisPhePheAspGlyAspGlyMetValHisAlaVal 180
Db 2311 CTTTCAGAGCCGGTGACAGGTCACCACTTCTTCAGCGAGACGGTATGTTGTTCCGCGTC 2370
Qy 181 LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal 200
Db 2371 AAATTCGAACACGGTTCAGTACGTAGCTAGCTTGGCGGTTTACTTCAGACTAACCGGTTT 2430
Qy 201 GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHis 220
Db 2431 CAGGAACGTCATATGGTCGACCGGTTTTTCCCAAAGCATCGGTGAGCTTTCACGGCCAC 2490
Qy 221 ThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaGlyIleValAspPro 240

```

Db 2491 ACCGGTATTCCCGGACTCATGCTATTCTACGCCAGAGCTGCAGCCGGTATAGTCGACCG 2550  
Qy 241 AlahisgIyThrGlyValAlaAenAlaGlyLeuValTyrPheAsnGlyValGluLeuAla 260  
Db 2551 GCACACGAACCGGTGAGCTACCGCGGTTTGGTCTATTTCATATGCGCGTTATTGGCT 2610  
Qy 261 MetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThr 280  
Db 2611 ATGTCGGAGGATGATTTACCTTACCAAGTTCAGATCACTCCCAATGGAGATTAAACAC 2670  
Qy 281 ValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysVal 300  
Db 2671 GTTGTCGGTTCGATTTTGTAGTGAATTTAGATTCACATGATTCACACCGCCGAAAGTC 2730  
Qy 301 AspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeu 320  
Db 2731 GACCGGAATCCGGTGAACCTTCGCTTTAAGCTACGACGCTGCTTTCAAGCCTTACCTA 2790  
Qy 321 LysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAsp 340  
Db 2791 AAATACCTCCGATTTCTACCGGACGGAATTAATCACCGGACGTCGAGATTCAGCTTGAT 2850  
Qy 341 GlnProThrMetHisAspPheAlaIleThrGluAsnPheValValValProAspGln 360  
Db 2851 CAGCCACGATGATGCAGATTTCCGCGATTACAGAGAACTTCGTCGCTACCTGACCA 2910  
Qy 361 GlnValValPheLysLeuProGluMetIleArgGlySerProValValTyrAspLys 380  
Db 2911 CAAGTCGCTTTCAAGCTCCCGAGATGATCCGGTGGGTCTCCGGTGGTTTACGACAAG 2970  
Qy 381 AsnLysValAlaAaArgPheGlyIleLeuAspLysTyrAlaGluAspSerAsnIleLys 400  
Db 2971 AACAGGTCCGAATTCGGGATTTTAGACAATACCGGAGATTCATCGACATTAAG 3030  
Qy 401 TrpIleAspAlaProAspCysPheCysPheHisLeuTrpAsnAlaTrpGluGluProGlu 420  
Db 3031 TGGATTGATGCTCCAGATGCTTCTGCTTCATCTCTGGAACGCTTGGGAAGAGCCAGAA 3090  
Qy 421 ThrAspGluValValIleGlySerCysMetThrProProAspSerIlePheAsnGlu 440  
Db 3091 ACAGATGAAGTCGTCGATAGGCTCTGTATGATCTCCACAGACTCAATTTTCAACGAG 3150  
Qy 441 SerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGlu 460  
Db 3151 TCTCAGCAGAAATCTCAAGAGTGTCTCTGTAATTCGCGCTGAATCTCAAAACCGGTAA 3210  
Qy 461 SerThrArgArgProIleIleSerAsnGluAspGlnValAsnLeuGluAlaGlyMet 480  
Db 3211 TCRACTCCGCTCCGATCATCTCAACGAAGATCAACAGTCAACCTCGAAGCAGGGATG 3270  
Qy 481 ValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluPro 500  
Db 3271 GTCACAGAAACATGCTCGGCGGTAAACCAAAATTCGCTTACTTGGCTTTAGCCGAGCCG 3330  
Qy 501 TrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHis 520  
Db 3331 TGGCCTAAAGTCTCAGGATTCGCTAAAGTTGATCTCACTCTGGAAGATTAAAGAACAT 3390  
Qy 521 LeuTyrGlyAspAsnArgTyrGlyGluProLeuPheLeuProGlyGlyGlyGluGlu 540  
Db 3391 CTTTACGGCGATACCGTTACGGAGGAGGCTCTGTTCTCCCGGAGGAGGAGAGAG 3450  
Qy 541 GluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLysSerGluLeu 560  
Db 3451 GAAGACGAAGATACATCTCTGTTTCTGTTCCACGACGAGAGACATCGAAATCGGAGTTA 3510  
Qy 561 GlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArgVal 580  
Db 3511 CAGATAGTTAAGCGCTTAGCTTAGAGGTTGACACACCGTTAACTTCGCTCAAGGTT 3570  
Qy 581 ProTyrGlyPheHisGlyThrPheIleGlyAlaAspLeuAlaLysGlnValVal 599

Db 3571 CCGTACGATTACGGTACATTCATCGGAGCCGATGATTGGCGAAGACGTCGTG 3627  
RESULT 4  
AB028617/c  
LOCUS AB028617 52232 bp DNA linear PLN 27-DEC-2000  
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, Pl clone:MOA2.  
ACCESSION AB028617 BA000014  
VERSION AB028617.1 GI:5041970  
KEYWORDS  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
1 (sites)  
Sato,S., Nakamura,Y., Kaneko,T., Kato,T., Asamizu,E. and Tabata,S.  
Structural analysis of Arabidopsis thaliana chromosome 3. I.  
Sequence features of the regions of 4,504,864 bp covered by sixty  
Pl and TAC clones  
DNA Res. 7 (2), 131-135 (2000)  
20277480  
10819329  
2 (bases 1 to 52232)  
REFERENCE  
Sato,S., Nakamura,Y., Kaneko,T., Kato,T., Asamizu,E. and Tabata,S.  
Direct Submission  
Submitted (09-JUN-1999) Yasukazu Nakamura, Kazusa DNA Research  
Institute, Department of Plant Gene Research; 1532-3, Yana,  
Kisarazu, Chiba 292-0812, Japan [E-mail:ynakamu@kazusa.or.jp,  
Tel:81-438-52-3935, Fax:81-438-52-3934]  
Address for correspondence: kaos@kazusa.or.jp  
For the latest information on annotation of this clone, please see  
http://www.kazusa.or.jp/kaos/cgi-bin/agd/graph.cgi?cc=MOA2  
Genes with similarity to proteins in the databases are described in  
'product' or 'note' qualifiers. Genes that have no significant  
protein similarity are described as 'unknown protein'.  
The software programs used to predict genes include: Grail  
(Informatics Group, Oak Ridge National Laboratory,  
http://compbio.ornl.gov/Grail-1.3/),  
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),  
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of  
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and  
SplicePredictor (Volker Brendel, Stanford University,  
http://grmmln1.cool.iastate.edu/cgi-bin/sp.cgi).  
Genes encoding tRNAs are predicted by tRNAscan-SE  
(Sean Eddy, Washington University School of Medicine, St. Louis,  
http://genome.wustl.edu/eddy/tRNAscan-SE/).  
This sequence may not be the entire insert of this clone. It may be  
shorter because we remove overlaps between neighboring submissions.  
The 5' clone is MLN21 and the 3' clone is MIE1.

FEATURES  
source

1..52232  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strains="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="3"  
/clone="MOA2"  
/clone\_lib="Mitsui Pl"  
Join(1996..2132,2225..2349,2455..2501,2648..2725,  
2807..2931,3021..3087,3216..3278,3370..3439,3529..3629,  
3718..3822,3906..4091)  
/note="gene id:MOA2.2"  
/codon\_start=1  
/evidence=not\_experimental  
/product="glycolate oxidase"  
/protein\_id="BAB01334.1"  
/db\_xref="GI:11994212"

## CDS

/translation="MEITNTEYDAIAKQKLPMVYDYASGADQWTLQNRNFAFR  
ILFRPLIDVSKIDMTTVLFGKISMPVAPTMQKVAHPDGEYATARAASAAGTI  
MTLSMAVSSVEEVAATGPGIRFQLYVVKRVVQVLEAQAQKAIATLTVDTPR  
LGRSDINKRFTLPNLTAKNFEGLDQKNDENDSLASVYAGQIDRTLSWKDVON  
IQRTIKLPILVKGVLTGEDARIATCAGAAGLIVSNHGQRQDYPATISALEEVVKAT  
QGRIPFLDGGVRRGTDFKALALGASGIFGRPVYFSLAAEGEAGVKVLQMLRDF







Satou, M., Seki, M., Shinn, P., Southwick, A., Tracy, S.E.,  
 Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.  
 Direct Submission  
 Submitted (12-SEP-2001) Plant Gene Expression Center, 800 Buchanan  
 Street, Albany, CA 94710, USA  
 RIKEN Genomic Sciences Center (GSC) members carried out the  
 collection and clustering of RAFL cDNAs (RAFL cDNA: "RIKEN  
 Arabidopsis Full-Length cDNA"): Seki, M., Narusaka, M., Ishida, J.,  
 Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,  
 Hayashizaki, Y. and Shinozaki, K.  
 The Salk, Stanford, PGSC (SSP) Consortium members carried out the  
 sequencing and annotation of the RAFL cDNAs: Yamada, K., Banno, J.,  
 Banno, F., Dale, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S.,  
 Quach, H.L., Tang, C.C., Toriumi, M., Yamamura, Y., Yu, G., Yu, S.,  
 Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C.,  
 Kossena, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Nguyen, M.,  
 Palm, C.J., Shinn, P., Southwick, A., Tracy, S.E., Davis, R.W.,  
 Ecker, J.R. and Theologis, A.  
 Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to  
 this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC)  
 contributed equally to this work as PIs.  
 Annotation is based on the January 2002 version of the Arabidopsis  
 Genome submitted to GenBank.  
 Location/Qualifiers  
 1..2331  
 /organism="Arabidopsis thaliana"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:3702"  
 /chromosome="3"  
 /clone="RAFL08-11-H16 (R11094)"  
 /note="This clone is in a modified pBluescript vector  
 (FLC-1) as a BamHI/XhoI insert."  
 ecotype: Columbia  
 1..2331  
 /gene="At3g14440"  
 1..122  
 /gene="At3g14440"  
 123..1922  
 /gene="At3g14440"  
 /codon\_start=1  
 /evidence="experimental"  
 /product="putative 9-cis-epoxycarotenoid dioxygenase"  
 /protein\_id="RAFL07104.1"  
 /db\_xref="GI:15810433"  
 /translation="WASETATAAVSGRWLGNHTQPPPLSSQSSDLYSCSLPMSRV  
 TKLVSSALHTFPALHPFKSSNSFALVVPKAKESNTKQNNLQFRAAAALDRAG  
 FVVSHEKHLPLKPTADPSVQIAGNFAPYNEQPVVRNLFPVVKGLPDSIKGVYVRGANGP  
 LHEPVTHHFFDQGMVHAKVFEHGSASYACRFOTNRFPVQERGLGRFPVFKAI GELH  
 GHTGIALRLMFLYARAAAGIVDPAGHTGVANAGLVYFNGRLLAMSEDDLPYQVQITPNG  
 DIKTVGRDPDGLSMTIAHPKVPESGELFALSVDVVSFKPLKYPRFSPDQKSPD  
 VRIQDQPMHDPALTENFVVPDQVQVFKLPEMILGGSFVVDKKNVAREGILDKY  
 ADSLNKWDPCFCFHLNMANEPEITDEVVIGSCMTFPDSIFNEDSNLKSVL  
 EIRLNKGTSTRRPIIISNDOQNLKAGMVRNMLRKRTKFAYLALAEAPKPVSGFA  
 KYDLTTEGVKHELYGDNRYGGEPLFLPQEGGEDEGYILCFVHDEKTKWSELQIVNAV  
 SLEVTATVKLPISRVYPGHGTPIGADLAKQVV"  
 misc\_difference 1234  
 /gene="At3g14440"  
 /note="compared to genomic sequence resulting in an amino  
 acid sequence difference"  
 /replaces="g"  
 1923..2331  
 /gene="At3g14440"  
 misc\_difference 2059  
 /gene="At3g14440"  
 /note="not present in genomic sequence"  
 misc\_difference 2312  
 /gene="At3g14440"  
 /note="not present in genomic sequence"  
 misc\_difference 2315  
 /gene="At3g14440"  
 /note="not present in genomic sequence"  
 /gene="At3g14440"

BASE COUNT 615 a 560 c 533 g 623 t  
 ORIGIN  
 /note="not present in genomic sequence"

Alignment Scores: 3,14e-219 Length: 2331  
 Pred. No.: 3143.00 Matches: 598  
 Percent Similarity: 99.83% Conservative: 0  
 Best Local Similarity: 99.83% Mismatches: 1  
 Query Match: 99.78% Indels: 0  
 DB: 8 Gaps: 0  
 US-09-758-269-6 (1-599) x AY056255 (1-2331)  
 QY 1 MetaLaserPheThrAlaThraAlaValSerGlyArgTTPLeuGlyGlyAsnHisThr 20  
 Db 123 ATGGCTTCTTTCACGCGAACCGCTCGGTTCTTGGAGATGGCTTGGGCAATCACTACT 182  
 QY 21 GlnProLeuSerSerSerGlnSerSerAspLeuSerTyrCysSerSerLeuProMet 40  
 Db 183 CAGCGCCCATTTATCGTCTTCTCAAGCTCCGACTTGGTATTGTAGCTCTTACCTATG 242  
 QY 41 AlaSerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProProAlaLeu 60  
 Db 243 GCAGTCGTGTCCACAGTAAGCTCAATGTTTCATCTGCGCTTCACACTCTCCAGCTCTT 302  
 QY 61 HisPheProLysGlnSerSerAsnSerProAlaLeuValValLysProLysAlaLysGlu 80  
 Db 303 CATTTCCCTTACCAATCATCAAACTCTCCGCCATTGTGTTAAGCCCAAGCCAAAGAA 362  
 QY 81 SerAsnThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAlaAla 100  
 Db 363 TCCACACATAACACAGATGAATTTGTTCCAGAGAGCGGCGGCGAGCGTTGGACCGCGCG 422  
 QY 101 GluGlyPheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSer 120  
 Db 423 GAGGGTTTCTTGTTCAGCCACGAGAGCTACACCCGCTTCTTAAACCGGTGATCTAGT 482  
 QY 121 ValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgArgAsnLeuPro 140  
 Db 483 GTTCAGATCGCGGAATTTTGTCTCGGTGAATGAACAGCCGCTCGGCGTAACTTCGCG 542  
 QY 141 ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnPro 160  
 Db 543 GTGGTCCGAAACCTTCCCGATTCCCATCAAGAGAGTGTATGTGCGACACGAGCTAACCA 602  
 QY 161 LeuHisGluProValThrGlyHisHisPheAspGlyAspGlyMetValHisAlaVal 180  
 Db 603 CTTACAGAGCGGTGACAGGTCAACCACTTCTTCGACGAGAGCGGTATGGTTTCAACGCGTC 662  
 QY 181 LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal 200  
 Db 663 AAATTCGAACACGGTTTCACTAGCTACGTTGCGGTTTACTCAGACTAACCGGTTTGT 722  
 QY 201 GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHis 220  
 Db 723 CAGAACGTCATTTGGTTCGACCGGTTTCCCAAGCCATCGGTGAGCTTCAACGCCAC 782  
 QY 221 ThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaGlyIleValAspPro 240  
 Db 783 ACCGGTATTCGCCGACTCATGCTATTCTACGCGAGAGCTGCAGCCGCTATGTCGACCGCG 842  
 QY 241 AlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAla 260  
 Db 843 GCACACGAACCGGTGTAGCTAACGCGGTTTGGTCTATTTCATGCGCGTTATTGGCT 902  
 QY 261 MetSerGluAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThr 280  
 Db 903 ATGTCCGAGGATGATTTACCTTACCAAGTTTCAGATCACTCCCAATGAGGATTTAAAAACC 962  
 QY 281 ValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysVal 300  
 Db 963 GTTGTGCGTTCGATTTTGTGACACATTTAGATCCCAATGATTTGCCACCCGCAAGTC 1022



Qy	110	LeuHisProLeuProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAlaPro	129
Db	412	GAGCACCCCTTTGCGGAAAAACAGCCGACCCAGGAGTCCAGATTTCCTGGGAATTTCCGCTCCG	471
Qy	130	ValAsnGluGlnProValArgArgAsnLeuP-ovalValValGlyLeuLeuProAspSerIle	149
Db	472	GTACCGGAAAAACCCAGTTTGCCAATCTCTTCGGTCACCGGAAAAAATACCCAAATGTGT	531
Qy	150	LysGlyValTyrValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHisHis	169
Db	532	CAAGGGGTTTACGTTTCGAACGGAGCTAACCTCTTTTCGAACCAACACCGCCGAGCCCAT	591
Qy	170	PhePheAspGlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerTyr	189
Db	592	TTCTTCGACGGCAGCGATGTTGCTACGCGGTTCAAATTCAAAAATGGGTACGCTAGTTAC	651
Qy	190	AlaCysArgPheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgProVal	209
Db	652	GCITGCGGTTCTACTGAAACAGAGAGTTTGTTCAGAAAAAAGCTTTGGGTGCCTGTT	711
Qy	210	PheProLysAlaIleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPhe	229
Db	712	TTCCCTAAAGCCATTGGTGAATTACATGCTGCTGGAATTCGAAGGCTTATGTGTGTTT	771
Qy	230	TyrAlaArgAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAlaSer	249
Db	772	TACGCTCGTGGGCTCTTCGGACATTATTGATCAGATGAGGAACTGGAGTTGCAAAACGCC	831
Qy	250	GlyLeuValTyrPheAsnGlyArgLeuLeuLeuAlaMetSerGluAspLeuProTyrGln	269
Db	832	GGTTTAGTCTATTCAATAACCCGATTACTTGTCTGATGCTGAAAGATGATTGTGCTTACCAT	891
Qy	270	ValGlnIleThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGln	289
Db	892	GTAAGGTAAACCCACCCGCGGATCTTAAACAGAGGGTCGATCGATTTCGACGCCACG	951
Qy	290	LeuGluSerThrMetIleAlaHisProLysValAspProGluSerGlyGluLeuPheAla	309
Db	952	CTAAATCCCAATGATGATGCTCACCAAAAGCTGCACCCAGTTTCCGGTGAGCTATTGTCT	1011
Qy	310	LeuSerTyrAspValValSerLysProTyrLeuLysTyrPheArgPheSerProAspGly	329
Db	1012	CTTAGCTACGATGTGATTTCAGAACCATCTCAGTATTTCAGATTTCAAAAAATGGG	1071
Qy	330	ThrLysSerProAspValGluIleGlnLeuAspGlnProThrMetMetHisAspPheAla	349
Db	1072	GAATAAATCAATGATGTGAAATCCAGTTGAAGACCCCAACATGATGCGATGATTTCGCA	1131
Qy	350	IleThrGluAsnPheValValProAspGlnGlnValValPheLysLeuProGluMet	369
Db	1132	ATTACTGAGAAATTCGTCATCATTCCTGATCAACCAAGTCGTTTCAAGATGCTCGAAATG	1191
Qy	370	IleArgGlyGlySerProValValTyrAspLysAsnLysValAlaArgPheGlyIleLeu	389
Db	1192	ATCCGTGGAGGTTACCCGGTGGTTTACGACAAAGAACCAAGTGTCCCGATTGGGATCTCG	1251
Qy	390	AspLysTyrAlaGluAspSerSerAsnIleLysTyrIleAspAlaProAspCysPheCys	409
Db	1252	GATAAGTACCGGAAAGATGGGTCTGATTGAAATGGGTGAAGTTCCTGATTGTTCTGT	1311
Qy	410	PheHisLeuThrAsnAlaTyrPgluGluProGluThrAspGluValValIleGlySer	429
Db	1312	TTCCATCTCTGGAATGCTTGGGAAGAACACAGATGAAATCTGTTGTAAATGGGTCA	1371
Qy	430	CysMetThrProProAspSerIlePheAsnGluSerAspGluAsnLeuLysSerValLeu	449
Db	1372	TGTATGACACCACACAGACTCCATTTCATGAATGATGAAGGACTAAAAAGCGTTTAA	1431
Qy	450	SerGluIleArgLeuAsnLeuLysThrGlyGluSerThrArgArgProIleIleSerAsn	469
Db	1432	TCCGAAATTCGCTCAATTTGAAACACAGGAAAAACACAGAAAGACCAATGAAATAC	1491
Qy	470	GluAspGlnGlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLys	489

Db	1492	CCGATGACACAGTGATTTAGAGGCTGGAATGGTGAACAGAAACAACTTTGGAAGAAA	1551
QY	490	ThrLysPheIaYrIeuAlaLeuAlaGluProTTPProLysValSerGlyPheAlaIys	509
Db	1552	ACACAGATGCTTATTGGCTATTGCTGAACCATGCCCAAGATTCTCTGGTTTGCAAAA	1611
QY	510	VallAspLeuThrThrGlyGluValLysLysHisLeuYrIGlyAspAsnArgTy-GlyGly	529
Db	1612	GTAGACCTGTTACAGGTGAAGTTGAGAATTCATTTATGTTGACACAAATATGGTGG	1671
QY	530	GluProLeuPheLeuProGlyGlu-----GlyGlyGluGluAspGluGlyIleLeu	547
Db	1672	GAGCGCTCTTTTTTACCACAGAGACCCAAACACAGCAGGAGGAGAGATGGTTATATT	1731
QY	548	CysPheValHisAspGluLysThrTTPLysSerGluLeuGlnIleValAsnAlaValSer	567
Db	1732	GCTTTCGTTTCATGATGAGAAAGATGGACATCAGAGCTTCAATTTGTTAATGCAATG	1791
QY	568	LeuGluValGluAlaThrValLysLeuProSerArgValProTyGlyPheHisGlyThr	587
Db	1792	TTAAAGTTAGAGGCAACAGTGAACACTTCCATCAAGAGTTCCTTATGGTTTTCATGGA	1851
QY	588	PheIleGlyAlaAspAspLeuAlaLysGln	597
Db	1852	TTCTAAATGCAATGATTTGGCAATCAG	1881
RESULT 7			
LENEOXANT		2171 bp	linear
LOCUS			PLN 22-MAY-1998
DEFINITION			Lycopersicon esculentum mRNA for nine-cis-epoxycarotenoid dioxygenase.
ACCESSION			Z97215
VERSION			Z97215.1 GI:2769641
KEYWORDS			nine-cis-epoxycarotenoid dioxygenase.
SOURCE			Lycopersicon esculentum (tomato)
ORGANISM			Lycopersicon esculentum
REFERENCE			Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
AUTHORS			Burbridge, A., Grieve, T.M., Jackson, A., Thompson, A. and Taylor, I.B.
TITLE			Structure and expression of a cDNA encoding a putative neoxanthin cleavage enzyme (NCE) isolated from a wilt-related tomato (Lycopersicon esculentum Mill.) library
JOURNAL			J. Exp. Bot. 47, 2111-2112 (1997)
REFERENCE			2. Burbridge, A. Direct Submission Submitted (01-JUL-1997) Burbridge A., The University of Nottingham, Physiology and Environmental Science, Sutton Bonington Campus, Loughborough, Leicestershire, LE12 5RD, UK
AUTHORS			Revised by [3]
TITLE			3 (bases 1 to 2171)
JOURNAL			Burbridge, A. Direct Submission Submitted (09-JAN-1998) Burbridge A., The University of Nottingham, Physiology and Environmental Science, Sutton Bonington Campus, Loughborough, Leicestershire, LE12 5RD, UK
REMARK			On Jan 13, 1998 this sequence version replaced gi:2243153.
AUTHORS			Location/Qualifiers
TITLE			1. 2171
JOURNAL			/organism="Lycopersicon esculentum"
FEATURES			/mol_type="mRNA"
source			/db_xref="taxon:4081"
			79.1896
			/codon_start=1
			/product="nine-cis-epoxycarotenoid dioxygenase"
			/protein_id="CAB10168.1"
			/db_xref="GI:2769642"
			/db_xref="GeneBank:U024002"
CDS			

NTNSSI.OAPPTLHFEPKOSNYOTPKNNTISHPKOENNNSSSSSTSCKWNLVOKAAAMAI





RESULT 8  
LES439079 19018 bp DNA linear PLN 13-MAR-2003  
LOCUS  
DEFINITION Lycopersicon esculentum nced1 gene for 9-cis-epoxycarotenoid  
dioxxygenase.  
ACCESSION  
VERSION A0439079.2 GI:28974076  
KEYWORDS 9-cis-epoxycarotenoid dioxxygenase; nced1 gene.  
SOURCE Lycopersicon esculentum (tomato)  
ORGANISM  
REFERENCE  
AUTHORS Thompson, A.J.  
TITLE Complementation of notabilis, a tomato mutant deficient in the  
abscisic acid biosynthetic enzyme 9-cis-epoxycarotenoid dioxxygenase  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 19018)  
AUTHORS Thompson, A.J.  
JOURNAL Direct Submission  
TITLE Submitted (13-MAR-2002) Thompson A.J., Plant Genetics and  
Biotechnology, Horticulture Research International, Wellesbourne,  
Warwick, CV35 9EF, UNITED KINGDOM  
REMARK revised by author [14-MAY-2002]  
COMMENT Related mRNA entry Z97215.  
FEATURES  
source  
1..19018  
/organism="Lycopersicon esculentum"  
/mol\_type="genomic DNA"  
/cultiivar="Money-maker"  
/db\_xref="taxon:4081"  
/chromosome="7"  
/map="7q40"  
/clone="NCEd1.1"  
267..567  
/rpt\_family="tomato anionic peroxidase inverted repeat  
(TAPTR)"  
repeat\_region  
/rpt\_type="INVERTED"  
2343..3038  
/rpt\_type="DIRECT"  
5304..7570  
/gene="nced1"  
5304..5311  
/gene="nced1"  
5334..5337  
/gene="nced1"  
5404..7570  
/gene="nced1"  
5482..7299  
/gene="nced1"  
/function="abscisic acid biosynthesis"  
/codon\_start=1  
/product="9-cis-epoxycarotenoid dioxxygenase"  
/protein\_id="CAD30202.1"  
/db\_xref="GI:28974077"  
/translation="MATTTSHATNTWIKTLSPSSKEFGFASNSISLLKNQHNROSL  
NINSLOAPLILHPKQSNVOTPKNNTISHPKQNNNNSSSSKWNLVKAAAMAL  
DAVESALTKLEHLEPKTADPRVOISGNFAPVNPVQCQLPVTKIPKVCQGVYR  
NGANPLFPTAGHHEFDGDMVHVOFKNGSACVACRETFETELVQKALGRPVYRKA  
IGELGHGIAKMLFYARGLFGLVDHSGKGVANAGLVYFNRLLANSEDDLLPHVK  
VPTGDLTEFGDFDGLKSTMIAPKLPDVSGLFALSIDVIQPKLYKTFRFSKNG  
EKNSDVEIPVEDPTWHDFAITENFVWIPDOQVFRKSEMIKGGSPVYDKNKVRFG  
ILDKYKDGSDLKWVEVDFCFHLNNAWEAEATDEIVVIGSCMTFPDSIFNECDEGL  
KVLSEIRLNLTGKSTRKSIENDEQVNLKAGVNRNKLGRKTEYAVLAIAEPWPK  
VSGFAKNLFTCEVEKFTYGDKNKGGELFLPRDPNSKEEDDGVILAFVHDEKEWKE  
LQIVNMSLKLEATVKLPSPVPYFGHTFINLANDLANQA"  
7570  
/gene="nced1"  
/evidence="experimental"  
8285..8575  
repeat\_region  
/rpt\_family="tomato anionic peroxidase inverted repeat  
(TAPTR)"

repeat\_region 10047..10285  
/note="putative"  
/transposon="stowaway element"  
10047..10063  
/rpt\_type="INVERTED"  
/rpt\_type="TERMINAL"  
10359..10756  
/note="putative mobile element"  
BASE COUNT 6823 a 2631 c 2645 g 6919 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 2.48e-155 Length: 19018  
Score: 2282.50 Matches: 436  
Percent Similarity: 81.83% Conservative: 64  
Best Local Similarity: 71.36% Mismatches: 92  
Query Match: 72.48% Indels: 19  
DB: 8 Gaps: 6  
US-09-758-269-6 (1-599) x LES439079 (1-19018)  
QY 1 MetAlaSerPheThrAlaThrAlaAlaValSerGlyArgTrpLeuGlyGlyAsnHisThr 20  
Db 5476 GTAGCTATGGCACTACTACTTCATCGCACAAATACATCGATT----- 5520  
QY 21 GlnProProLeuSerSerSerGlnSerSerAspLeuSerTyrCysSer-----Ser 37  
Db 5521 AAGACTAAGTTGTCATCGCATCATCAAGAGGTTTGGTTTTCGATCAAACTCTATTCT 5580  
QY 38 LeuProMetAlaSerArgValThrArgLysLeuValSerSerAlaLeuHisThrPro 57  
Db 5581 CTACTCAAAATCAACATAATAGGCAAGTCTCAACATTAATCTCTCTCTCAAGCTCA 5640  
QY 58 ProAlaLeuHisPheProLysGlnSerSerAsn-----SerProAla-----IleVal 73  
Db 5641 CCTATACTTCAATTTCTTAAACAATCTTCAAAATTAACAACACCAAGATAATAACAAT 5700  
QY 74 ValLysProLysAlaLysGlnSerAsn-----ThrLysGlnMetAsnLeu 88  
Db 5701 TCACACCCAAAACAGAAAACCAACACTCTCTCTTCTTCAACTTCCAAGTGGATTTA 5760  
QY 89 PheGlnArgAlaAlaAlaAlaLeuAspAlaAlaGluGlyPheLeuValSerHisGlu 108  
Db 5761 GTGCAGAAAGCAGCAGCAATGCTTTAGATGCTGTAGAAAGTCTTTAACTAAACATGAA 5820  
QY 109 LysLeuHisProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAla 128  
Db 5821 CTTGAACACCTTTGCCGAAAACAGCCGACCCGAGTCCAGATTCTTGGGAATTTGCT 5880  
QY 129 ProValAsnGluGlnProValArgAsnLeuProValValGlyLysLeuProAspSer 148  
Db 5881 CCGGTACCGGAAAATCCAGTCTGTCAATCTCTTCGGGTACCGGAAAATATCCCAATGT 5940  
QY 149 IleLysGlyValTyrValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHis 168  
Db 5941 GTTCAAGGCGTTTACGTTTGAACGAGGAGTAACTCTCTTTTGAACCAACCCCGGACAC 6000  
QY 169 HisPhePheAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSer 188  
Db 6001 CATTTCTCGACGCGACGGTATGGTTTACGCGGTTCAATTCAAAAATGGTGGCTAGT 6060  
QY 189 TyrAlaCysArgPheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgPro 208  
Db 6061 TAGCGTTGCGGTTTCTACTGAAACAGAGAGCTTTTCAAGAAAACCTTTGGTGCCCT 6120  
QY 209 ValPheProLysAlaIleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeu 228  
Db 6121 GTTTTCCCTAAACCCATTGGTGAATTCATGGTCACTCTGGAATTCGAAGGCTTATGCTG 6180  
QY 229 PheTyrAlaArgAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsn 248  
Db 6181 TTTTACGCTCGTGGGCTCTTCGGACTTGTGTGATCAGACAGTAAAGAACTGGTGTGCAAC 6240



```
QY 249 AlaGlyLeuValTyrPheAsnGlyArgLeuLeuAlaMetSerClnuAspLeuProTyr 268
Db 6241 GCGGGTTTGTCTATTTCATTAACCCGATTCTTGTCTATCTCGAAGATGTTTGCCTTAC 6300
QY 269 GlnValGlnIleThrProAsnGlyAspLeuLeuValThrValGlyArgPheAspGly 288
Db 6301 CATGTAAGGTAACACCCACCGCGATCTTAAACACAGAGGTCGATTTCGATTCGACGCG 6360
QY 289 GlnLeuGlnSerThrMetIleAlaHisProLysValAspProGlnSerGlyGluLeuPhe 308
Db 6361 CACCTAAATCCACCATGATAGTCTACCCAAAGCTCGACCCGATTTCCCGTGAGCTATT 6420
QY 309 AlaLeuSerTyrAspValValSerLysProTyrLeuLeuValThrValGlyArgPheSerProAsp 328
Db 6421 GCTCTTACGTACGATGATGATTCAGAACGCCATACCTCAAGTACTCTTCAGATTTTCAAAAAAT 6480
QY 329 GlyThrLysSerProAspValGluIleGlnLeuAspGlnProThrMetMetHisAspPhe 348
Db 6481 GGGGAAAAATCAAAATGATGTTGAAATTCACGTTGAAAGACCCACCAATGATGATGATTC 6540
QY 349 AlaIleThrGluAsnPheValValProAspGlnGlnValValPheLysLeuProGlu 368
Db 6541 GCATTTACTGAGAACTTCGTCGTCATCTCTGATCAACAGTCGTTTCAAGATGTCGAA 6600
QY 369 MetIleArgGlyGlySerProValValTyrAspLysAsnLysValAlaArgPheGlyIle 388
Db 6601 ATGATCCGTGGAGGTTCCACCGTGGTTTACGACAAGAACAAAGTTTCCCGATTGTTGATT 6660
QY 389 LeuAspLysTyrAlaGluAspSerSerAsnIleLysTrpIleAspAlaProAspCysPhe 408
Db 6661 CTGGATTAAGTACCGGAAAGATGGTCTGATTTGAAATGGTTGAAGTACTGATTTGTTTC 6720
QY 409 CysPheHisLeuTrpAsnAlaTrpGluGluProGluThrAspGluValValIleGly 428
Db 6721 TGTTTCCACCTCTGCAATGCTTGGGAAAGACGACAAACAGATGAAATCGTTGTAATTGT 6780
QY 429 SerCysMetThrProProAspSerIlePheAsnGlnSerAspGluAsnLeuLysSerVal 448
Db 6781 TCATGTATGACACCCACGACTCCATTTTCAATGAATGCTGTAAGGGCTAAAGATGTT 6840
QY 449 LeuSerGluIleArgLeuAsnLeuLeuThrGlyLeuSerThrArgArgProIleIleSer 468
Db 6841 TTATCCGAAATCCGTCCTCAATTTGAAACAGGGAATCAACAGAAATCCATAATCGAA 6900
QY 469 AsnGluAspGlnGlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArg 488
Db 6901 AACCCGGATGAACCAAGTGAAATTAGAAGCTGGAATGTTGAACCGAACCAAACTCGGAAG 6960
QY 489 LysThrLysPheAlaTyrLeuAlaLeuAlaGluProTrpProLysValSerGlyPheAla 508
Db 6961 AAAACAGAGTATGCTTATTTGGCTATCGCTGAACCATGGCCAAAGTTTCTGTTTTCGA 7020
QY 509 LysValAspLeuThrThrGlyGluValLysHisLeuTyrGlyAspAsnArgTyrGly 528
Db 7021 AAAATAAACCTGTTTCCCGGTGAAGTTGAGAAATTCATTATGTTGACAAACAATATGTT 7080
QY 529 GlyGluProLeuPheLeuProGlyGlu-----GlyGlyGluGluAspGluGlyTyrIle 546
Db 7081 GGGGAACCTCTTTTTTACCAGAGACCCACAGACGAGGAGGAGGACGATGTTTATATT 7140
QY 547 LeuCysPheValHisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaVal 566
Db 7141 TTAGCTTTCGTTTCAGATGAGAAAGAAATGGAATTCAGAACTGCAAAATGTTTAAACGCAATG 7200
QY 567 SerLeuGluValGluAlaThrValLysLeuProSerArgValProTyrGlyPheHisGly 586
Db 7201 AGTTTGAAAGTTGGAGGCACCTGTAAGCTTCCATCAAGAGTTCCTTATGATGTTTCATGGA 7260
QY 587 ThrPheIleGlyAlaAspAspLeuAlaLysGln 597
Db 7261 ACAITTCATAACCGCAATGATTTGGCAATCAG 7293
```

```
RESULT 9
AX148316
LOCUS 1818 bp DNA linear PAT 08-JUN-2001
DEFINITION Sequence 15 from Patent EP1116794.
ACCESSION AX148316
VERSION AX148316.1 GI:14347203
KEYWORDS
SOURCE
ORGANISM
Lycopersicon esculentum (tomato)
Lycopersicon esculentum
Lycopersicon esculentum
AUTHORS Iuchi,S., Kobayashi,M. and Shinozaki,K.
TITLE Transgenic plants carrying neoxanthin cleavage enzyme gene
JOURNAL Patent: Bp 1116794-A 15 18-JUL-2001;
Riken (JP)
FEATURES
source
1..1818
/organism="Lycopersicon esculentum"
/mol_type="genomic DNA"
/db_xref="taxon:4081"
CDS
1..1818
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC41199.1"
/db_xref="GI:14347204"
/translation="MATTSHANTWTKLSPSSKEFGFASNSILLKNQNRQSL
NINSLQAPILHPKQSSNYQTPKNTIISHPKENNSSSSSTKWNLYQKAAVAL
DAVSALTKEHEHPKTPKAPVQISGNFAPVFNPCQSLPVTGIPKRCQGVYVR
NGANPLPEPTAGHHFFDGMGVAHQKNGSAYACRFETRELVLQEKALGRPVYVR
IGELHSGIARLMLFYARGLGLVADHSKGTGVANAGLVVFNRLLAMSDDLPLPHVK
VTPGDLTKSRGPDFDGLSKMTAHSLKLPVSGELFALSDVIQKPLKYRFSKNG
EKNDVIPEDPTMMDPAITENFVVIPOQVVFKNSEMRGSPVYVKNKYSRG
ILDYAKGDGSLRWVEVDFCFHLNMAEEADELVIWGSCTPDSIFNEDEGL
KSVLSERLMLTKGTRKSIENPDQVNLBAGMNRNKLGRTEYAYLAISPWRK
VSGFAKVLFTGEVKFIYGDNGGGEPLFPRDPSKEDDDGVLAVHDEKWKSE
LQIVNMSLKLEATVKLPSRPVYGFHGTFFNANDLANQA"
BASE COUNT 569 a 383 c 369 g 497 t
ORIGIN
Alignment Scores:
Pred. No.: 1,24e-156 Length: 1818
Score: 2280.50 Matches: 435
Percent Similarity: 82.04% Conservative: 63
Best Local Similarity: 71.66% Mismatches: 90
Query Match: 72.40% Indels: 19
DB: Gaps: 6
US-09-758-269-6 (1-599) x AX148316 (1-1818)
QY 5 ThrAlaThrAlaAlaValSerGlyArgTrpLeuGlyGlyAsnHisThrGlnProLeu 24
Db 7 ACTACTACTTCCATGCCCAAAATACATGGATT-----AAGACTAAGTTG 51
QY 25 SerSerSerGlnSerSerAspLeuSerTyrCysSer-----SerLeuProMetAla 41
Db 52 TCAATGCCCATCATCAAGAGGTTTGGTTTTCATCAAACTCTATTCTCTACTCAAAAT 111
QY 42 SerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProAlaLeuHis 61
Db 112 CCAATAATAGGCAAGTCTCAACATTAAATTCCTCTCTTCAAGCTCCACCTATATCTCAT 171
QY 62 PheProLysGlnSerSerAsn-----SerProLala-----IleValValLysProLys 77
Db 172 TTTCTTAACAATCTTCAAAATATCAAAACACCAAGAAATAATAACAATTCACCCCAAAA 231
QY 78 AlaLysGluSerAsn-----ThrLysGlnMetAsnLeuPheGlnArgAla 92
Db 232 CAAGAAACACCAACTCTCTTCTTCTTCAACTTCCAAAGTGGAAATTTAGTCGAGAAACA 291
QY 93 AlaAlaAlaLeuAspAlaAlaGluGlyPheLeuValSerHisGluLysLeuHisPro 112
```

Db 292 GCAGCAATGGCTTAGATGCTGAGAAAGTGGCTTTAACTAAACATGAACCTTGAACACCT 351  
Qy 113 LeuProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAlaProValAsnGlu 132  
Db 352 TTCCGAAAAACAGCCAGCCAGGACGAGTTCCTGGGAATTTCTCCCGGTACCGGAA 411  
Qy 133 GlnProValArgAsnLeuProValValGlyLysLeuProAspSerIleLysGlyVal 152  
Db 412 AATCCAGTCTGTCAATCTCTTCGGTCCCGTACCGGAAAAATACCCAAATGTGTCAAGCGTT 471  
Qy 153 TyrValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHisShePheAsp 172  
Db 472 TAGCTTCGAACGAGCTAAACCTCTTTTGAACCAACCGCGACACCACTTTCTTCGAC 531  
Qy 173 GlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerValAlaCysArg 192  
Db 532 GGCACCGGTATGGTTCAGCGCTTCAATCARAAATGGTGGCTAGTACGCTTCCCGT 591  
Qy 193 PheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPheProLys 212  
Db 592 TTCACTGAACAGAGAGCTTGTTCAGAAAAAGCTTTGGGTGCGCTGTTTCCCTAAA 651  
Qy 213 AlaIleGlyLysLeuHisGlyHisThrGlyLysAlaArgLeuMetLeuPheTyrAlaArg 232  
Db 652 GCCATTGGTGAATACATGTCATCTGGAATTCGAAGGCTATGCTGTTTACGCTCGT 711  
Qy 233 AlaAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAlaGlyLeuVal 252  
Db 712 GGGCTCTCGGACTGTGTGATCACAGTACAGTAAAGAACTGGTGTGCAACCGCGGTTAGTC 771  
Qy 253 TyrPheAsnGlyArgLeuLeuAlaMetSerGluAspAspLeuProTyrGlnValGlnIle 272  
Db 772 TATTTCATTAACCGATTAATCTGCTATGCTGAAGATGATTTGCTTACCATGTAAGGTA 831  
Qy 273 ThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeuGluSer 292  
Db 832 ACACCCACCGGCGATCTTAAACAGAGGGTGCATTCGATTTCCGACGCCAGCTAAATCC 891  
Qy 293 ThrMetIleAlaHisProLysValAspProGluSerGlyGluLeuPheAlaLeuSerTyr 312  
Db 892 ACCATGATAGTCAACCAAGCTCGACCACTTCCGGTGGCTATTTGCTTCTAGTAC 951  
Qy 313 AspValValSerLysProTyrLeuLysTyrPheArgPheSerProAspGlyThrLysSer 332  
Db 952 GATGTGATTCAGAACCCATACCTCAAGTACTTCAGATTTTCAAAAAATGGGAAAAATCA 1011  
Qy 333 ProAspValGluIleGlnLeuAspGlnProThrMetHisAspPheAlaIleThrGlu 352  
Db 1012 AATGATGTTGAAATCCAGTTCAGAGCCCAACAATGATGATGATTTTCGCAATTAAGTAC 1071  
Qy 353 AsnPheValValProAspGlnGlnValValPheLysLeuProGluMetIleArgGly 372  
Db 1072 AACTTCGTCCTGCTCATCTGATCAACAGTCTGTTTCAAGATGCTGTAATGATCCGTGA 1131  
Qy 373 GlySerProValValTyrAspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyr 392  
Db 1132 GGTTCACCGGTGGTTTACGACAGAACAAAGTTTCCCGATTTGGTATTCTCGATAAGTAC 1191  
Qy 393 AlaGluAspSerSerAsnIleLysTrpIleAspAlaProAspCysPheCysPheHisLeu 412  
Db 1192 GCGAAAGATGGGTCTGATTTGAATGGTGGTGAAGTACCTGATGTTTCTGTTTCCACCTC 1251  
Qy 413 TrpAsnAlaTrpGluGluProGluThrAspGluValValIleGlySerCysMetThr 432  
Db 1252 TGGAAATGTTGGGAAAGACAGAAACAGATGAAATCGTTGTAATTTGGTTTCATGATGACA 1311  
Qy 433 ProProAspSerIlePheAsnGluSerAspGluAsnLeuLysSerValIleSerGluIle 452  
Db 1312 CCACCAAGATCCATTTTCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1371  
Qy 453 ArgLeuAsnLeuLysThrGlyGluSerThrArgTyrProIleIleSerAsnGluAspGln 472  
Db 1372 CGTCTCATTTGAACAGGGAATCAACAGAAATCCATAATCGAAACCCGGATGAA 1431

Qy 473 GlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPhe 492  
Db 1432 CAACTGATTTTAGAAGCTGGAATGGTGAACCGAAACAACTCGGAAGGAAACAGAGTAT 1491  
Qy 493 AlaTyrLeuAlaLeuAlaGluProTyrProLysValSerGlyPheAlaLysValAspLeu 512  
Db 1492 GCTTATTGGCTATCGCTGAACCATGCGCAAAAGTTTCTGGTTTGCAAAAGTAAACCTG 1551  
Qy 513 ThrThrGlyGluValLysLysHisLeuTyrGlyAspAsnArgTyrGlyGlyGluProLeu 532  
Db 1552 TTCACCGGTGAAGTTGAAATTCATTTATGGTGACCAACAATATGTTGGGAACCTCTT 1611  
Qy 533 PheLeuProGlyGlu-----GlyGlyGluGluAspGluGlyTyrIleLeuCysPheVal 550  
Db 1612 TTTTACAGAGAGCCCAACAGCAAGAGAGACGATGCTTATATTTAGCTTTTCGTT 1671  
Qy 551 HisAspGlyLysThrTrpLysSerGluLeuGlnIleValAsnAlaValSerLeuGluVal 570  
Db 1672 CACGATGAAAGATGGAATCGAAATCGAAATTTGTAACGCAATGAGTTTGAAGTTG 1731  
Qy 571 GluAlaThrValLysLeuProSerArgValProTyrGlyPheHisGlyThrPheIleGly 590  
Db 1732 GAGGCAACTGTGAAGCTTCCATCAAGATTCCTTATGATTTTCATGGAACATTCATAAC 1791  
Qy 591 AlaAspAspLeuAlaLysGln 597  
Db 1792 GCCAATGATTTGGCAATCAG 1812

## RESULT 10

BD017436 1818 bp DNA linear PAT 27-AUG-2002  
LOCUS Transgenic plant using neozanthine cleaving enzyme gene.  
DEFINITION  
BD017436  
ACCESSION  
BD017436.1 GI:22558612  
VERSION JP 2001258579-A/8.  
KEYWORDS  
Lycopersicon esculentum (tomato)  
SOURCE  
Lycopersicon esculentum  
ORGANISM  
Lycopersicon esculentum (tomato)  
REFERENCE  
1. (bases 1 to 1818)  
Iuchi, K., Kobayashi, M. and Shinozaki, K.  
AUTHORS  
Transgenic plant using neozanthine cleaving enzyme gene  
TITLE  
Patent: JP 2001258579-A 8 25-SEP-2001;  
JOURNAL THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH  
COMMENT  
OS Lycopersicon esculentum (tomato)  
PN JP 2001258579-A/8  
PD 25-SEP-2001  
PF 11-JAN-2001 JP 2001003476  
PI KIYOSHI IUCHI, MASATOMO KOBAYASHI, KAZUO SHINOZAKI PC  
C12N15/09,A01H5/00,C12N5/10,C12N9/02,C12N15/00,C12N5/00 CC  
Transgenic plant using neozanthine cleaving enzyme gene PH

## FEATURES (1).. (1818).

FT CDS Location/Qualifiers  
source  
1. .1818  
/organism="Lycopersicon esculentum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4081"  
BASE COUNT 569 a 369 g 497 t  
ORIGIN

## Alignment Scores:

Pred. No.: 1.24e-156 Length: 1818  
Score: 2280.50 Matches: 435  
Percent Similarity: 82.04% Conservative: 63  
Best Local Similarity: 71.66% Mismatches: 90  
Query Match: 72.40% Indels: 19  
Gaps: 6

US-09-758-269-6 (1-599) x BD017436 (1-1818)

QY 5 ThrAlaThrAlaValSerGlyArgTTPLeuGlyGlyAsnHisThrGlnProProLeu 24  
DB 7 ACTACTACTTACATGCCACAAATACATGATT-----AAGACTAAGTTG 51  
QY 25 SerSerSerGlnSerSerAspLeuSerTyrCysSer-----SerLeuProMetAla 41  
DB 52 TCAATGCCATCATCAAGAGGTTTGGTTTGGATCAAACTCTATTTCTCTACTCAAAAAT 111  
QY 42 SerArgValThrArgLysLeuAsnValSerSerAlaLeuHisThrProProAlaLeuHis 61  
DB 112 CAACATAATAGGCAGCAAGCTCTCAACATTAATCTCTCTTCAAGCTCCACTACTTCAAT 171  
QY 62 PheProLysGlnSerSerAsn-----SerProAla-----IleValValLysProLys 77  
DB 172 TTTCTCTAAACAATCTTCAAAATTAACAACACCAAGAAATTAATAAATTTACACCCCAAAA 231  
QY 78 AlaLysGlnSerAsn-----ThrLysGlnMetAsnLeuPheGlnArgAla 92  
DB 232 CAAGAAAACACACTCTCTTCTTCTCAACTTCCAGTGGATTTAGTCGAGAAAGCA 291  
QY 93 AlaAlaAlaLeuAspAlaAlaGluGlyPheLeuValSerHisGluLysLeuHisPro 112  
DB 292 GCAGCAATGCTTTAGATGCTGTAGAAAGTGTCTTAACTAAACATGAACCTTGAAACACCT 351  
QY 113 LeuProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAlaProValAsnGlu 132  
DB 352 TTGCGGAAAACGCCGACCCACGAGTCCAGATTTCTGGGAATTTTGTCTCGGTACCGGAA 411  
QY 133 GlnProValArgAsnLeuProValValGlyLysLeuProAspSerLysGlyVal 152  
DB 412 AATCCAGTGTCTCAATCTTTCGGTCCCGGAAAAATACCAAAATGTGTCAAGCGCTT 471  
QY 153 TyrValArgGlnGlyAlaAsnProLeuHisGlnProValThrGlyHisPhePheAsp 172  
DB 472 TACGTTCGAACCGGAGTAACTCTTTTGAACCAACCCCGGACACCACTTTCTTCGAC 531  
QY 173 GlyAspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerTyrAlaCysArg 192  
DB 532 GGCAGCGGTATGTTTCAGCGGTTCAATTCAAAAATGGGTGGCTAGTTACGCTTGGCGT 591  
QY 193 PheThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPheProLys 212  
DB 592 TTCTACTGAAACAGAGAGGCTTGTTCAGAAAAAGCTTTGGGTGCGCTGTTTTCCTCTAAA 651  
QY 213 AlaIleGlyGluLeuHisGlyHisThrGlyLeuAlaArgLeuMetLeuPheTyrAlaArg 232  
DB 652 GCCATTGGTGAATTCATGTTGCTACTCTGGAATTCAGAGGCTTATGCTGTTTTACGCTCGT 711  
QY 233 AlaAlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsnAlaGlyLeuVal 252  
DB 712 GGGCTCTTCGGACTTGTGATCACAGTAAAGGAACCTGTTGCAACGCGCGTTTAGTC 771  
QY 253 TyrPheAsnGlyArgLeuLeuAlaMetSerGluAspAspLeuProTyrGlnValGlnIle 272  
DB 772 TATTTCAATACCGATTAATCTTCTGCTATGCTGAAGATGATTTGCTTACCATGTAAGGTA 831  
QY 273 ThrProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeuGluSer 292  
DB 832 ACACCCACCGCGCATCTTAAACACAGAGGCTGATTCGATTTTCGACGCCAGCTAAATCC 891  
QY 293 ThrMetIleAlaHisProLysValAspProGluSerGlyGluLeuPheAlaLeuSerTyr 312  
DB 892 ACCATGATAGTCTACCCCAAACTGACCCAGCTTTCCGGTGAGCTATTTGCTCTTAGTAC 951  
QY 313 AspValValSerLysProTyrLeuLysTyrPheArgPheSerProAspGlyThrLysSer 332  
DB 952 GATGTGATTCAGAGCCATACCTCAAGTACTTCAGATTTTCAAAAAATGGGAAAAATCA 1011  
QY 333 ProAspValGluIleGlnLeuAspGlnProThrMetHisAspPheAlaIleThrGlu 352  
DB 1012 AATGATGTTGAATTCAGATGGAAGACCAACATGATGCAATTTTCGCAATTAATCTGAG 1071  
QY 353 AsnPheValValProAspGlnGlnValValPheLysLeuProGluMetIleArgGly 372

DB 1072 AACTTCGTCGTCATCTCTGATCAACAAGTCGTTTTCAGATGTCGTGAATGATCCCGGGA 1131  
QY 373 GlySerProValValTyrAspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyr 392  
DB 1132 GGTTCACCGTGGTTTACGACAGAAAGATTTCCCGATTTGGTATTTCTGGATAAGTAC 1191  
QY 393 AlaGluAspSerSerAsnIleLysTrpIleAspAlaProAspCysPheCysPheHisLeu 412  
DB 1192 CGCAAGAGATGGTCTGATTTGAATGGTTGAAGTACCTGATTTCTGTTTCCACCTC 1251  
QY 413 TrpAsnAlaTrpGluGluProGluThrAspGluValValIleGlySerCysMetThr 432  
DB 1252 TGGATGCTTGGGAAGACGAAACAGATCAATCGTTGTAATGTTCTCATGTATGACA 1311  
QY 433 ProProAspSerIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIle 452  
DB 1312 CCACGAGACTCCATTTTCAATGAATGTGATGAAGGGCTAAAGAGTGTTTTATCCGAATC 1371  
QY 453 ArgLeuAsnLeuLysThrGlyGluSerThrArgArgProIleIleSerAsnGluAspGln 472  
DB 1372 CGTCTCAATTTGAACAGGGAATCAACAGAAAAATCCATATCGAAAAACCCGATGAA 1431  
QY 473 GlnValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPhe 492  
DB 1432 CAAGTGAATTTAGAGCTGGAATGTGAACCGAAACAACTCGGAAGAAAAACAGAGTAT 1491  
QY 493 AlaTyrIleuAlaLeuAlaGluProTrpProLysValSerGlyPheAlaLysValAspLeu 512  
DB 1492 GCTTATTTGGCTATCGCTGAACCATGGCCAAAAGTTTCTGGTTTTGCAAAAGTAAACCTG 1551  
QY 513 ThrThrGlyGluValLysLysHisLeuTyrGlyAspAsnArgTyrGlyGlyProLeu 532  
DB 1552 TTCACCGGTGAAGTTGAGAAATTCATTTATGTTGACAACAATATGTTGGGAAACCTCTT 1611  
QY 533 PheLeuProGlyGlu-----GlyGlyGluLysGluGlyTyrIleLeuCysPheVal 550  
DB 1612 TTTTACCAGAGAGACCCCAACAGCAAGGAAGACGATGTTATATTTAGCTTTCTGTT 1671  
QY 551 HisAspGluLysThrTrpLysSerGluLeuGlnIleValAsnAlaValSerLeuGluVal 570  
DB 1672 CACGATGAGAAAGATGGAATCAGAACTGCAATTTGTAACGCAATGAGTTGAAGTTG 1731  
QY 571 GluAlaThrValLysLeuProSerArgValProTyrGlyPheHisGlyThrPheIleGly 590  
DB 1732 GAGGCAACTGTGAAGCTTCCATCAAGAGTCTCTTATGGAATTCATGGAACATTATTAAC 1791  
QY 591 AlaAspAspLeuAlaLysGln 597  
DB 1792 GCCAATGATTGGCAAAATCAG 1812

## RESULT 11

AF190462

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

1 (bases 1 to 2398)

Qin, X. and Zeevaert, J. A.

The 9-cis-epoxycarotenoid cleavage reaction is the key regulatory

step of abscisic acid biosynthesis in water-stressed bean

Proc. Natl. Acad. Sci. U.S.A. 96 (26), 15354-15361 (-1999)

20079657

10611386

REFERENCE

2 (bases 1 to 2398)

Phaseolus vulgaris 9-cis-epoxycarotenoid dioxygenase (NCED1) mRNA, complete cds.

AF190462.1 GI:6715256

Phaseolus vulgaris

Phaseolus vulgaris

Bakayota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

Phaseolus.

1 (bases 1 to 2398)

Qin, X. and Zeevaert, J. A.

The 9-cis-epoxycarotenoid cleavage reaction is the key regulatory

step of abscisic acid biosynthesis in water-stressed bean

Proc. Natl. Acad. Sci. U.S.A. 96 (26), 15354-15361 (-1999)

20079657

10611386

REFERENCE

2 (bases 1 to 2398)

AUTHORS Qin, X. and Zeevaert, J.A.D.  
TITLE Direct Submission  
JOURNAL Submitted (27-SEP-1999) MSU-DOE Plant Research Laboratory, Michigan State University, Wilson Str., East Lansing, MI 48824, USA  
FEATURES Location/Qualifiers  
source 1..2398  
/organism="Phaseolus vulgaris"  
/mol\_type="mRNA"  
/cultivar="Top Crop"  
/db\_xref="taxon:3885"  
gene 1..2398  
/gene="NCED1"  
CDS 124..1971  
/gene="NCED1"  
/function="catalyzes the oxidative cleavage of the C-11,12 double bond of 9-cis-epoxycarotenoids (C40) to yield xanthoxin (C15) and C25-apo-aldehydes"  
/codon\_start=1  
/product="9-cis-epoxycarotenoid dioxygenase"  
/protein\_id="AAF26356.1"  
/db\_xref="GI:6715257"  
/translation="MPSPASNTWINTTLPSSCSPPFKDLASTSSPTLLPFRKSSSSNTNITCSGLQTLRPOYQFISSTSTTTTPPIRPTKTDPAKVQIAGNPAFVPEHAADQALPVQKPKIDGIVYVANGANPLYEFVAGHFFDGDGMHAKFTNGAASYACRFETQRLAQEKSLRPFPPKAIAGLHSGIARLLFYARSLFQIVDGHGMVYANAGLYTFNNHLLMSBDDLPYHVAITSGNGLTTVGRYDFNGLNSTIAHPKLDPVQVDFLHLSYDVVQKPLYKFRSADGVKSPDVEIPLKPTWMDPAITENFVVPDQVFPKLTWMTGGSPVYDKNKSRSFGLIKNAKDANAMWIDAPCFCLANAMEEPTDEIVIGSCMTPASINFECDSEKSVLSIRLNLRTGKSTRPIISDAEQVNLKAGMUNRKLGRKTQFALYALAEKPFKVGAKVDLFPSEGVKMYGEKFGGEFLFPLNGEERGGYILAFVHDEKWKSELIQVNAQNLKLEASIKLPSVPVYGFHGTFIHSDLRKQA"

BASE COUNT 627 a 687 c 516 g 568 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,448-148 Length: 2398  
Score: 2172.00 Matches: 428  
Percent Similarity: 78.21% Conservative: 60  
Best Local Similarity: 68.59% Mismatches: 100  
Query Match: 68.95% Indels: 36  
Dbs: 8 Gaps: 9  
US-09-758-269-6 (1-599) x AF190462 (1-2398)

QY 3 SerPheThrAlaThraAlaAlaValSerGlyArgTrpLeuGlyGlyAsnHisThrGlnPro 22  
Db 115 TCAATAAGCATGCTTCCACAGCTTCAAATACCTTGATT-----AACACCACACTCCCA 168  
QY 23 ProLeuSerSerGlnSerSerAspLeuSerTyrCysSerSer----- 37  
Db 169 TCTTCTGTCTTCCCTTTAAAGACCTGCTCCACATTTTCTCTCCACACCTTA 228  
QY 38 LeuProMetAlaSerArgValThrArgLysLeuAsn---ValSerSerAlaLeuHisThr 56  
Db 229 CTTCCTTCAAGAAAGATCTCTTCCACACCAACACCATCATCATGTTCTCTTCAACA 288  
QY 57 ProProAlaLeuHisPheProLysGln-----SerSerAsnSerPro 70  
Db 289 -----CTCATCTACCCCAACAGTACCAACCAACATCCATCCACACCAACCAACC 339  
QY 71 AlaIleValValLysProLysAlaLysGlu----- 80  
Db 340 CCAACTCCGATTAAACCCACCACCAACCAACCAACCAACCAACCAACCAACCAACCA 399  
QY 81 -----SerAsnThrLysGln-----MetAsnLeuPheGlnArgAla 93  
Db 400 CTTCTCTCCGACACCAACCAACCAATTTCTCTCAAAATGGAACCTTTCTCGAAGAGCCCT 459  
QY 94 AlaAlaAlaLeuAspAlaAlaGluGlyPheLeuValSerHisGlyLysLeuHisProLeu 113  
Db 460 GCGACAGGCTTGATATGTTGGAGACGGCGCTGTTTCGCACGAGAGCAACACCGCTC 519

QY 114 ProLysThrAlaAspProSerValGlnIleAlaGlyAsnPheAlaProValAsnGluGln 133  
Db 520 CCCAAAACGGCGGACCCGAAAGTCCAAATCCCGGGAACCTTCGCGCGGTCCCGAGCAC 579  
QY 134 ProValArgArgAsnLeuProValValGlyLysLeuProAspSerIleLysGlyValTyr 153  
Db 580 GCGCGGATCAAGCCCTTCGGTGGTGGAAAAATCCCAAAATGCAATTCAGCGGTGTAC 639  
QY 154 ValArgAsnGlyAlaAsnProLeuHisGluProValThrGlyHisHisPhePheAspGly 173  
Db 640 GTGCGCAACGGCGGGAATCCCGCTCTACGAGCCGTTCGCGGGGACCACTTCTTCGACGC 699  
QY 174 AspGlyMetValHisAlaValLysPheGluHisGlySerAlaSerTyrAlaCysArgPhe 193  
Db 700 GAGCGCATGTCACCGCTGCAAGTTCACGAACGGCGGTGCAGTACGCTCGCGCTTC 759  
QY 194 ThrGlnThrAsnArgPheValGlnGluArgGlnLeuGlyArgProValPheProLysAla 213  
Db 760 ACCGAACACAGCGTCTCGCGAAGAGAAATCTTTAGGCCCGCGGTGTTCCCAAGGCT 819  
QY 214 IleGlyGluLeuHisGlyHisThrGlyIleAlaArgLeuMetLeuPheTyrAlaAla 233  
Db 820 ATCGCGGAACCTCCACGGCCACTCCGGCATAGCTCGCTTCTCTTCTACGCTCGCAGC 879  
QY 234 AlaAlaGlyIleValAspProAlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyr 253  
Db 880 CTCCTCCAGCTGTTGATGGGTCCACGGCATGGGGTGGGGAACCGCGGTCTCGTCTAT 939  
QY 254 PheAsnGlyArgLeuLeuAlaMetSerGluAspAspLeuProTyrGlnValGlnIleThr 273  
Db 940 TTCAACACCACTCTTGGCCATGTCGAGAGAGATTTTACCATACCACTGAGAATACCC 999  
QY 274 ProAsnGlyAspLeuLysThrValGlyArgPheAspPheAspGlyGlnLeuGluSerThr 293  
Db 1000 TCTAACGGCGACTTAACCACTGTCGGCTTACGACTTCAACGGCGCACTCAACTCCACA 1059  
QY 294 MetIleAlaHisProLysValAspProGluSerGlyGluLeuPheAlaLeuSerTyrAsp 313  
Db 1060 ATGATCGCCACCCGAAACTCGACCCCTCAACGGCGACTTCCACGGCTCAGCTACGAC 1119  
QY 314 ValValSerLysProTyrLeuLysTyrPheArgPheSerProAspGlyThrLysSerPro 333  
Db 1120 GTCGTTTCAGAGACCTTACCTCAAGTACTTCCGTTTCTCCGTGACGGGTAAAGTCCGCC 1179  
QY 334 AspValGluIleGlnLeuAspGlnProThrMetHisAspPheAlaIleThrGluAsn 353  
Db 1180 GAGCTCGAATCCCTGAAAGAGCCACCATGATGACGATTTTCGCAATACCGAGAT 1239  
QY 354 PheValValProAspGlnValValPheLysLeuProGluMetIleArgGlyGly 373  
Db 1240 TTGCTGCTGCTCCCGACCCAGCAGGTGGTCTTCAAACCTAACAGAGATGATCACCGCGGC 1299  
QY 374 SerProValValTyrAspLysAsnLysValAlaArgPheGlyIleLeuAspLysTyrAla 393  
Db 1300 TCCCCGCTGCTTACGACAAAGAACAAACCTCACCGCTTTGGGATTTTGGATAAGATGCG 1359  
QY 394 GluAspSerSerAsnIleLysTrpIleAspAlaProAspCysPheCysPheHisLeuTrp 413  
Db 1360 AAGGACCGAATGCGATGCGGTGGATCGACGGCGCGGAGTGTCTGTGTTCCATCTCTGG 1419  
QY 414 AsnAlaTrpGluProGluThrAspGluValValIleGlySerCysMetThrPro 433  
Db 1420 AACCGTGGGAGAGGCTGAAACCCAGAGATGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1479  
QY 434 ProAspSerIlePheAsnGluSerAspGluAsnLeuLysSerValLeuSerGluIleArg 453  
Db 1480 GCGGACTCTTCAATGATGAGAGTTCGAGAGCGGTTCGAGAGATGAGTGGTGGTGGTGG 1539  
QY 454 LeuAsnLeuLysThrGlyGluSerThrArgArgProIleIleSerAsnGluAspGln 473  
Db 1540 CTGAATTCAGGACGGGCAAGTCGATCGCCCGCCCATTTATCTCC---GACCCCAACAA 1596  
QY 474 ValAsnLeuGluAlaGlyMetValAsnArgAsnMetLeuGlyArgLysThrLysPheAla 493

```

Db      1597  GTGAACCTTGAAGCCGCGCATGTGTGAACAGAAACAAGCTAGGAAGAAACCTAGTCGCT 1656
Qy      494  TylLeuAlaLeuAlaGluProTrpProLysValSerGlyPheAlaLysValAspLeuThr 513
Db      1657  TATCTGGCTCTGGCCGAGCCCTGGCCCAAGTCTCGGGTTTTCGGAAGTGAATTGTTTC 1716
Qy      514  ThrGlyGluValLysLysHisLeuTyrGlyAspAsnArgTyrGlyGluProLeuPhe 533
Db      1717  AGTGGGGAAGTGCAAGATACATGTATGGAGAAGAAAGTTCGGTGGGAGGCTCTGTTT 1776
Qy      534  LeuProGlyGluGlyGluGluAspGluGlyTyrIleLeuCysPheValHisAspGlu 553
Db      1777  CTTCC---AACGCCGAAGAGGTGAAGGTATATCTGGCATCTGTGACAGCG 1833
Qy      554  LysThrTrpLysSerGluLeuGluValAlaValSerLeuGluValGluAlaThr 573
Db      1834  AAAGAGTGAATCCGAGCTGCAGATTGTGAATGCCCAAAATTTAAAGTCGAAGCTTCC 1893
Qy      574  ValLysLeuProSerArgValProTyrGlyPheHisGlyThrPheIleGlyAlaAsp 593
Db      1894  ATCAAACTCCCTCTCGCGTTCTTATGTTTTCATGGCATTCTTTATCTTAAGGAT 1953
Qy      594  LeuAlaLysGln 597
Db      1954  TTGAGGAAGCAA 1965

RESULT 12
LOCUS   AX652128 95769 bp DNA linear PAT 22-MAR-2003
DEFINITION Sequence 1979 from Patent WO03000898.
ACCESSION AX652128
VERSION   AX652128.1 GI:29154942
KEYWORDS
SOURCE   Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1
AUTHORS Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M.,
Karagiri, F., Qian, S., Tao, Y., Whitham, S., Xie, Z., Zhu, T. and Zou, G.
TITLE Plant genes involved in defense against pathogens
JOURNAL Patent: WO 03000898-A 1979 03-JAN-2003;
Syngenta Participations AG (CH)
FEATURES
source
1. 95769
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
BASE COUNT 30958 a 16961 c 16682 g 31168 t
ORIGIN

Alignment Scores:
Pred. No.: 3,45e-146 Length: 95769
Score: 2170.50 Matches: 419
Percent Similarity: 80.61% Conservative: 80
Best Local Similarity: 67.69% Mismatches: 87
Query Match: 66.90% Indels: 33
DB: 6 Gaps: 9

US-09-758-269-6 (1-599) x AX652128 (1-95769)
Qy      1  MetAlaSerPheThrAlaThrAlaAlaValSerGlyArgTrpLeu----- 15
Db      33716  ATGGCTTCTACTACTTACTTCTCTTCTACTTCAACCCCAATTCCTCGACCGAAGCTTCTCA 33775
Qy      16  --GlyGlyAsnHisThrGlnProLeuProLeuSerSerGlnSerSerAspLeuSerTyr 34
Db      33776  ACTTCATCTTCTCTCTCGAGCCTAAACTTCAATCT-----CTATCTTTT 33820
Qy      35  CysSerSerLeuPro-----MetAlaSerArgValThr 45
Db      33821  TCCTAACACATTAGAAACAAGAAACCTAGTCCCTGCTATGTTCTTCTCCGTCAC 33880
Qy      46  ArgLysLeuAsnValSerSerAlaLeuHisThr-----ProProAlaLeuHis 61
Db      33881  AAAAAATCTTCGCTCGTCTCTCTCTCCATCTCCGACATTCAAACACACCGCTTGG--- 33937
Qy      62  PheProLysGlnSerSerAsnSerProAlaIleValLysLysProLysAlaLysGluSer 81
Db      33938  -----AAGAAACATATTAACGATGTCCCAACTTGATC-----CCCAAAAACGACGAATCAG 33988
Qy      82  AsnThrLysGlnMetAsnLeuPheGlnArgAlaAlaAlaLeuAspAlaAlaGlu 101
Db      33989  AATCAAAA---TTAAACCCCGGTGCAGAAACCGCCCTATGGTTTATAGACGCGGTGGAA 34045
Qy      102  GlyPheLeuValSerHisGlu---LysLeuHisProLeuProLysThrAlaAspProSer 120
Db      34046  AACGCCATGATCTCACAGAGCGCCGCGCTCATCCCATCTCTAAACGGCGGATCTCTGCC 34105
Qy      121  ValGlnIleAlaGlyAsnPheAlaProValAsnGluGlnProValArgArgAsnLeuPro 140
Db      34106  GTCCAAATAGCCGGAACTTCTCCCGGTACCGGAGAAACCTGTCGTGCATAAACCCTCCG 34165
Qy      141  ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnPro 160
Db      34166  GTGACTGGAACAGTACCAGATGTCATTCAAGGAGTTTACGTGACAAACGGAGCAATCCA 34225
Qy      161  LeuHisGluProValThrGlyHisPhePheAspGlyAspGlyMetValHisAlaVal 180
Db      34226  CTTCAAAACAGTCTCCGGCCACCATTTGTTCCAGCGAGACGCTATGGTACACGCTGTC 34285
Qy      181  LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal 200
Db      34286  CGGTTCGATAACGGCTCGGTTAGCTACGCTTGTCCGTTTACCGAAACAAACCGGTTGGTT 34345
Qy      201  GlnGluArgGlnLeuGlyValGtrProValPheProLysAlaIleGlyGluLeuHisGlyHis 220
Db      34346  CAGAAGCTGTAATGGTGGTCCGCTTCGCCAAGCAATTCGAGAGCTTCACGGACAT 34405
Qy      221  ThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAspPro 240
Db      34406  TTAGAAATCGCAAGCTTATGCTCTTCAATACCCCGCGGCTATTGGGTTCAGTCCGCCG 34465
Qy      241  AlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAla 260
Db      34466  ACCGGAGACTCGGTGTCGCTAACCGCGTTAGTTATTTCAATGGTCATCTCTTAGCC 34525
Qy      261  MetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThr 280
Db      34526  ATGTCGGAAGACGATTTACCGTACCATGTCAAAGTTACTCAAACCGGAGATTTAGAAACT 34585
Qy      281  ValGlyArgPheAspPheAspGlyGlnLeuGluSerThrMetIleAlaHisProLysVal 300
Db      34586  TCGGGTCGGTACGATTTTCAGCGTCAGTTAAATCAACATGATAGCCACCCCGAANAATC 34645
Qy      301  AspProGluSerGlyGluLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeu 320
Db      34646  GATCCGAAACCCGAGAACTCTTCGCTTTAAGCTACGACGCTGTTCAAAGCTTACTTATA 34705
Qy      321  LysTyrPheArgPheSerProAspGlyThrLysSerProAspValGluIleGlnLeuAsp 340
Db      34706  AAATACTTCAGATTCAATCCGACGTCAGAAATCACCAGAGCTCGAATTCGCGTTGAT 34765
Qy      341  GlnProThrMetMetHisAspPheAlaIleThrGluAsnPheValValValProAspGln 360
Db      34766  CAACCAACGATGATCCAGATTTCCGATCACTAGATATTCGTAGTATTCAGACCAAA 34825
Qy      361  GlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyrAspLys 380
Db      34826  CAAGTGTATTCCGATTCGCGAGATGATAGAGGTGTTCTCCGCTGGTTCCTACACGAG 34885
Qy      381  AsnIleValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLys 400
Db      34886  AAGAAGAAATCAAGATTTCGGGATTTTGAATAAATAAAGCGGAAGATGCTTCGTCGATCAA 34945

```

```

Qy 401 TPIleaspAlapProaspCysPheCysPheHisLeuTrpAsnAlaTrpGluProGlu 420
Db 34946 TGGATCGAAGTACCAAGTTGTTTGTTCATCATATGACTCTTGGAGAACCCAGAA 35005
Qy 421 ThrAspGluValValValledGlySerCysMetThrProProAspSerIlePheAsnGlu 440
Db 35006 ACAGACGAGGTGTCTGTGATTCATGATCATGACGACCCACCTGATTCATATTTTCAACGAA 35065
Qy 441 SerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGlu 460
Db 35066 CACGACGAAACACTTCAGAGTGTGTGTGCGAGATAAGCTTAACCTTAACACGGGAA 35125
Qy 461 SerThrArgArgProIleSerAsnGluAspGlnGlnValAsnLeuGluAlaGlyMet 480
Db 35126 TCAACACGAGTACCGGTATCTCTG-----GAACAAGTTAATCTCGAAGCTGGATG 35176
Qy 481 ValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyLeuAlaLeuAlaGluPro 500
Db 35177 GTCAACCGGAATTTATTAGTAGAAACCCCGTATGCTTATCTGCTTTAACCGAACCG 35236
Qy 501 TrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHis 520
Db 35237 TGGCTTAAGTGTCCGGTTCGCTAAAGTGGACTTATCGACCGGAGAGATTCGAAGTAT 35296
Qy 521 LeuTyArgAspAsnArgTyArgGlyGluProLeuPheLeuProGlyGlyGlyGlu 540
Db 35297 ATTACCGGAGGAGGAAATACGAGGAGAGCCTCTGTTCTACCTTCGGTGACCGAGAA 35356
Qy 541 GluAspGluGlyTyLeuLeuCysPheValHisAspGluLysThrTrpLysSerGluLeu 560
Db 35357 GAAGACGAGGTACATAATATGTTGCTTCACGAGGAGGAAGTGAATTCGGAATC 35416
Qy 561 GlnIleValAsnAlaValSerLeuGluValGluAlaThrValLysLeuProSerArgVal 580
Db 35417 CAACCTATAACCGCGTTAATATGAAGCTTGAAGCTACCGTACCGCTTCGCTCGAGAGT 35476
Qy 581 ProTyGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGlnValVal 599
Db 35477 CTTACGTTTCCACGGAAGCTTATCAGTAGGAAGATTTATCGAAGCAAGCTTTG 35533

RESULT 13
AC013430
LOCUS
DEFINITION
Genomic sequence for Arabidopsis thaliana BAC F3F9 from chromosome
I, complete sequence.
AC013430
VERSION
AC013430.5 GI:8096768
KEYWORDS
HTG.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 95769)
Chao,Q., Brooks,S., Buehler,E., Johnson-Hopson,C., Khan,S., Kim,C.,
Shinn,P., Altafi,H., Bai,B., Choi,J., Conn,L., Conway,A., Hansen,N.,
Lee,J., Lenz,C., Li,J., Liu,A., Liu,K., Li,J., Li,J., Li,J., Li,J.,
Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A.,
Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Federspiel,N.,
Theologis,A., and Ecker,J.R.
Genomic sequence for Arabidopsis thaliana BAC F3F9 from chromosome
I.
Unpublished
2 (bases 1 to 95769)
Ecker,J.R.
Direct Submission
Submitted (11-NOV-1999) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
3 (bases 1 to 95769)
Ecker,J.R.

TITLE
JOURNAL
REFERENCE
AUTHORS
Direct Submission
Submitted (07-FEB-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
4 (bases 1 to 95769)
Chen,K., Shinn,P., Brooks,S., Buehler,E., Chao,Q.,
Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bai,B., Chin,C.,
Choi,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N.,
Hwang,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J.,
Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H.,
Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M.,
Yu,G., Davis,R., Federspiel,N., Theologis,A., and Ecker,J.
Direct Submission
Submitted (24-MAY-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
5 (bases 1 to 95769)
Chen,K., Shinn,P., Brooks,S., Buehler,E., Chao,Q.,
Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bai,B., Chin,C.,
Choi,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N.,
Hwang,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J.,
Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H.,
Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M.,
Yu,G., Davis,R., Federspiel,N., Theologis,A., and Ecker,J.
Direct Submission
Submitted (26-MAY-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
On May 26, 2000 this sequence version replaced gi:6921155.
COMMENT
FEATURES
source
1..95769
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F3F9"
join(71..402,574..649,732..932)
note="similar to oxidosqualene cyclase dbj|BAA33462.1"
/codon_start=1
/evidence=not_experimental
/product="F3F9.1"
/protein_id="AAF71792.1"
/db_xref="GI:8052528"
/translation="MGHVOLLTFCYSPFFSFFLFFLLFNKNNVILISVDITVGT
FELKAYEQITPPVKVEDGSGTHERMEYGEFFLSRLCRPPMTGLEKSLDLS
FLHNVSSFGHLEIEPDAHREELRYIYRLNDDGGLHVECKSMFCTALNYI
CLRIRSGDDEGYACKLISKMPWITVFLIYNNESCPSI"
complement(join(3438..3746,3868..4032,4110..4664))
note="similar to F-box domain gb|AAF14684.1"
/codon_start=1
/evidence=not_experimental
/product="F3F9.2"
/protein_id="AAF71813.1"
/db_xref="GI:8052549"
/translation="MLASWICSLTSMKTKNLIANPNRFAFTGDNVLKSTFTFN
APSUTLSYQILQDAQANNCKNLYLDLQIFAEVDVFNVEVISCSEVLILOI
IFFNPGHKLDKHDKILSNQNDISIEVRAARDILSIEYPCSDNVLEIPRL
QFNNVAGRLPHTSINTSCPPQESNGMKWDTNTATSPASLSLDKNPREV
EVLKILAVTMEKIEVIFSNPLGECNSDGCANNLMEKLEPPFNADPRVDT
IWMNFKGSNKEQFALASRFVQKTMVKMKWIKTSDEKKEKKEIAVAKLKELPGN
EELSTECF"
join(7138..7400,7491..7887)
note="unknown protein; similar to EST gb|AI999747.1"
/codon_start=1
/evidence=not_experimental
/product="F3F9.3"
/protein_id="AAF71793.1"
/db_xref="GI:8052529"
/translation="METGLNLIKSLCVLVVGSVAQAAPWNPNGRPGTCGDHYE
CPTVKLEAGYGFIRYDAALNTSPISLSMTQATKTGFERLRYLEGDNKSNVK
MMYAPIAQTATPGRSVYTSVLIPKKNQONPQADLHVRSKTYPIVAVRQICGVIS
NNVADEAAALMESILNWLPIEKSKGLPAYFLAVYNPSPHTTARVINELWPFN
M"

```











829	ATGTCGGAAGACGATTACCTTACCACGTCGAGAAATCACCCCTAAACGGCGCACTTTAAACCCACC	888
281	ValGlyArgPheAspPheAspGlyGlnLeuSerThrMetIleAlaHisProLysVal	300
889	GTGTGGCGGTTACGACTTCAACGGCGAGCTCAACTCAACAATGATCGCCACCCGAAACTG	948
301	AspProGluSerGlyGlnLeuPheAlaLeuSerTyrAspValValSerLysProTyrLeu	320
949	GACCCGTCGACGGCGACCTCCACGGCTCAGCTACGACTCATTCAGAAGCCCTTACCTC	1008
321	LysTyrPheArgPheSerProAspGlyThrLysSerProAspValGlnLeuGlnLeuAsp	340
1009	AAGTACTTCGGTTCTCCCGCGACGGGTCGAAGTCCCGACGTGGAAATCCCCCTGAAG	1068
341	GlnProThrMetMetHisAspPheAlaIleThrGlnAsnPheValValProAspGln	360
1069	GAGCCCAACCATGATGCACGATTTCCGCCATAACGGAGAATTCGTCTGCTGTCGCCGACG	1128
361	GlnValValPheLysLeuProGluMetIleArgGlyGlySerProValValTyrAspLys	380
1129	CAGGTGCTCTCAAACTTAACGGAGATGATCACCGCGGGTCCCCCGTGGTCTACGACAAG	1188
381	AsnLysValAlaArgPheGlyIleLeuAspLysTyrAlaGluAspSerSerAsnIleLys	400
1189	AACAAAACCTCACGGTTTGGGATCTCCACAAGAAATCGGAAGGACGGGAATCGCATGGG	1248
401	TrpIleAspAlaProAspCysPheCysPheHisLeuTrpAsnAlaTrpGluProGlu	420
1249	TGGATCAGCGCGCGGATGTGTTCTGTCTCCACTCTGGAACGGTGGGAGGACCGCGAA	1308
421	ThrAspGluValValIleGlySerCysMetThrProProAspSerIlePheAsnGlu	440
1309	ACCAGAGAGGTGTGGTGATTTGGGTCTGATGACCCCTGGGACTCCATTTTCAACGAA	1368
441	SerAspGluAsnLeuLysSerValLeuSerGluIleArgLeuAsnLeuLysThrGlyGlu	460
1369	TCGAGGAGAGTGTGAAGAGCGTCTGTACAGATAAAGCTTGAACCTTGAGACCGCGCAAG	1428
461	SerThrArgArgProIleIleSerAsnGluAspGlnValAsnLeuGluAlaGlyMet	480
1429	TCCACTCGGCGCGCCATTATCTCC---GACCGCGAACAAAGTGAACCTGGAACCGCGCATG	1488
481	ValAsnArgAsnMetLeuGlyArgLysThrLysPheAlaTyrLeuAlaLeuAlaGluPro	500
1486	GTGAACAGAAACAAGCTCGGAAGGAAGACCCAGTTCGGCTATCTGGCTCTGGCGAGGCC	1544
501	TrpProLysValSerGlyPheAlaLysValAspLeuThrThrGlyGluValLysLysHis	520
1546	TGGCCCAAAGTCTCGGGCTTTGCGAAAGTTGATTTCGTGAGTGGGAAGTGAAGAAGTAC	1600
521	LeuTyrGlyAspAsnArgTyrGlyGlyGluProLeuPheLeuProGlyGlyGlyGlu	540
1606	ATGATATGAGAAGAGAAGTTCGGTGGGGAGCCTCTGTTCTTCCC---AACGGCCAAAAA	1666
541	GluAspGluGlyTyrIleLeuCysPheValHisAspGluLysThrTrpLysSerGluLeu	560
1663	GAAGACGATGGGTATATTCTGGCATTCGTGCACGACGAGAAGAATGGAATTCGAGCTG	1720
561	GlnIleValAsnAlaValSerLeuGlnValGluAlaThrValLysLeuProSerArgVal	580
1723	CAGATTGTGAATGCCAAAATTTAAAGCTCGNAGCTTCCATCAAACTCCCTCTCGTGT	1780
581	ProTyrGlyPheHisGlyThrPheIleGlyAlaAspAspLeuAlaLysGln	597
1783	CCCTACGGTTTTCATGGAACTTTCATTCTCAAGGATTTTCAGAGAAACAA	1833

Search completed: November 16, 2003, 22:18:21  
Job time : 6294 secs

```

/mol_type="genomic DNA"
/db_xref="taxon:3917" 355 t
BASE COUNT 447 a 595 c 442 g
ORIGIN
Alignment Scores:
Pred. No.: 1.78e-148 Length: 1839
Score: 2168.50 Matches: 423
Percent Similarity: 78.44% Conservative: 61
Best Local Similarity: 68.56% Mismatches: 98
Query Match: 68.84% Indels: 35
DB: 6 Gaps: 8

US-09-758-269-6 (1-599) x BD017434 (1-1839)
Qy 8 AlaAlaValSerGlyAsgTtpLeuGlyGlyAsnHisThrGlnProProLeuSerSerSer 27
   : : : : :
Db 7 TCATCACTTCAAAACACTTGGTTTAAGCCACACTCCCACTCTCCCTCTCAAGACCTA 66

Qy 28 GlnSerAspLeuSerTyrCySerSerLeuProMetAlaSerArgValThrArgLys 47
   : : : : :
Db 67 CDTTCCACATCT---TCTCCACAAACTTACTTCTCTTA-----AGAAACATCTCT 117

Qy 48 LeuAsn---ValSerSerAlaLeuHisThrProAlaLeuHisPheProLysGln--- 65
   : : : : :
Db 118 TCCAAACACATCATGATGTTCCCTTCAACA-----CTCCACTTCCCAACAGTAC 168

Qy 66 -----SerSerAsnSerProAlaIleValValLysProLysAlaLysGlnSer 81
   : : : : :
Db 169 CAACCAACATCCACATCCACATCCACAGCCACCAACACACCCCAATCAAAACT 228

Qy 82 AsnThr----- 83
   : : : : :
Db 229 ACCACCATCACCAACACACACCGCCAGGAAACCAACCTCTCTCTGACACCAACAA 288

Qy 84 -----LysGlnMetAsnLeuPheGlnArgAlaAlaAlaAlaLeuAspAlaAla 100
   : : : : :
Db 289 CCATTACCTCAAAATGGAACCTTCTCCAGAAAGCCGCTGCCACGGCTTCGAGCTG 348

Qy 101 GluGlyPheLeuValSerHisGluLysLeuHisProLeuProLysThrAlaAspProSer 120
   : : : : :
Db 349 GAACGGCGCTCGTCTCGCAGCGGCAACACCGCTCTCCCAAAACGCGGACCCGAGG 408

Qy 121 ValGlnIleAlaGlyAsnPhaAlaProValAsnGluGlnProValArgAsnLeuPro 140
   : : : : :
Db 409 GTCCAAATCGCGGGAACCTTCGCGCGGTGCGGAGCATGCGCGGATCGAGATCTCCG 468

Qy 141 ValValGlyLysLeuProAspSerIleLysGlyValTyrValArgAsnGlyAlaAsnPro 160
   : : : : :
Db 489 GTGGTTCGAAAAATCCCAATGCATTGACGGCGGTACGTGCGCAACGGTSCCAATCG 528

Qy 161 LeuHisGluProValThrGlyHisHisPhePheAspGlyAspGlyMetValHisAlaVal 180
   : : : : :
Db 529 CTCCTACGAGCCTGTGCGCGGCAACCACTTCCTTCGACGGCGAGCATGGTCTCACCG 588

Qy 181 LysPheGluHisGlySerAlaSerTyrAlaCysArgPheThrGlnThrAsnArgPheVal 200
   : : : : :
Db 589 AAGTTCACGAACGGCGCCCGCAGCTACGCTTCGCGCTTCACCGAGACGACGCTCTCT 648

Qy 201 GlnGluArgGlnLeuGlyArgProValPheProLysAlaIleGlyGluLeuHisGlyHis 220
   : : : : :
Db 649 CAGGAGAAATCTCTAGCGCGCGCGGTGTTCCCGAAAGCCCATCGGGAGCTCTCACG 708

Qy 221 ThrGlyIleAlaArgLeuMetLeuPheTyrAlaArgAlaAlaAlaGlyIleValAspPro 240
   : : : : :
Db 709 TCCGGCATCGCGGGCTCTCTCTTCTACGGCGCGGCTCTCTTCGGGCTCGTTGATGG 768

Qy 241 AlaHisGlyThrGlyValAlaAsnAlaGlyLeuValTyrPheAsnGlyArgLeuLeuAla 260
   : : : : :
Db 769 TCCAGGGATCGGGCTGCGAAGCGCGGTCTGCTACTTCAACAAACCACTCTTTGGCC 828

Qy 261 MetSerGluAspAspLeuProTyrGlnValGlnIleThrProAsnGlyAspLeuLysThr 280
   : : : : :

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 21:17:43 ; Search time 36 Seconds  
(without alignments)

3037.587 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150

Sequence: 1 MASFTTAAGVGRWLGNGHNT.....VPYGFHGTFIGADDLAKQVV 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications, AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3150	100.0	599	10	US-09-758-269-6
2	2280.5	72.4	605	10	Sequence 6, Appl
3	2168.5	68.8	612	10	Sequence 16, Appl
4	1991	63.2	583	10	Sequence 2, Appl
5	1930	61.3	604	10	Sequence 14, Appl
6	1663.5	52.8	577	10	Sequence 10, Appl
7	968	30.7	595	10	Sequence 4, Appl
8	938	28.8	538	10	Sequence 8, Appl
9	937	23.7	538	10	Sequence 18, Appl
10	936	29.7	538	10	Sequence 33, Appl
11	411.5	13.1	505	8	Sequence 22, Appl
12	379	8.9	177	12	Sequence 56, Appl
13	260	8.3	526	15	Sequence 1, Appl
14	247	7.8	506	15	Sequence 4, Appl
15	238.5	7.6	549	12	Sequence 19, Appl

Sequence 17, Appl  
Sequence 21, Appl  
Sequence 5, Appl  
Sequence 2, Appl  
Sequence 83, Appl  
Sequence 77, Appl  
Sequence 7784, Appl  
Sequence 36, Appl  
Sequence 37319, A  
Sequence 32438, A  
Sequence 4, Appl  
Sequence 6, Appl  
Sequence 13187, A  
Sequence 135, App  
Sequence 5107, App  
Sequence 2, Appl  
Sequence 1119, App  
Sequence 252, App  
Sequence 428, App  
Sequence 21, Appl  
Sequence 43, Appl  
Sequence 126, App  
Sequence 54, Appl  
Sequence 110, App  
Sequence 2, Appl  
Sequence 58, Appl  
Sequence 3692, App  
Sequence 40, Appl

US-10-168-517-17  
US-10-168-517-21  
US-10-053-192-5  
US-10-168-517-2  
US-09-759-508B-2  
US-10-171-311-83  
US-09-322-261-77  
US-10-032-585-7784  
US-10-410-881-36  
US-09-864-761-37319  
US-10-029-386-32438  
US-10-203-351-4  
US-10-103-377C-6  
US-09-815-242-13187  
US-10-315-515-135  
US-09-815-242-5107  
US-10-251-503-2  
US-10-238-075-1119  
US-10-085-959-252  
US-09-934-455-428  
US-09-905-129-21  
US-09-991-630-21  
US-10-301-822-49  
US-10-032-189-126  
US-10-176-847-54  
US-10-177-293-110  
US-09-272-975-2  
US-09-272-975-58  
US-09-738-626-3692  
US-10-410-881-40

#### ALIGNMENTS

RESULT 1  
US-09-758-269-6  
; Sequence 6, Application US/09758269  
; Patent No. US20020104120A1  
; GENERAL INFORMATION:  
; APPLICANT: IUCHI, SATOSHI  
; APPLICANT: KOBAYASHI, MASATOMO  
; APPLICANT: SHINOZAKI, KAZUO  
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN  
; FILE REFERENCE: 3914-3  
; CURRENT APPLICATION NUMBER: US/09/758,269  
; PRIOR FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: JP 2001-003476  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: JP 2000-010056  
; PRIOR FILING DATE: 2000-01-13  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 599  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-09-758-269-6

Query Match 100.0%; Score 3150; DB 10; Length 599;  
Best Local Similarity 100.0%; Pred. No. 7.9e-309;  
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASFTTAAGVGRWLGNGHNTQPLSSQSSDLSYCSSLPMSGRVTRKLVSSALHTPPAL 60  
DB 1 MASFTTAAGVGRWLGNGHNTQPLSSQSSDLSYCSSLPMSGRVTRKLVSSALHTPPAL 60  
QY 61 HFPKQSSNPAIVVFKAKESNTKQNLFORAAAALDAAGFLVSHKHLPLKPTADPS 120  
DB 61 HFPKQSSNPAIVVFKAKESNTKQNLFORAAAALDAAGFLVSHKHLPLKPTADPS 120  
QY 121 VQIAGNAPVNEQVPRNLVVGVKLPDSIKGVVNGANPLPEPVTGHHFFDGDGWHAV 180

Db 121 VOIAGNAPVNEQVRRNL PVGKLPDSIKGVYVRNGANLPHBPVGTGHFFDGDGMVHAV 180  
Qy 181 KFEHGSASYACRFTQTNRVQEROLGRPVFPFKKAIIGELHGTGIALMLFYARAAAAGIVDP 240  
Db 181 KFEHGSASYACRFTQTNRVQEROLGRPVFPFKKAIIGELHGTGIALMLFYARAAAAGIVDP 240  
Qy 241 AHGTGVANAGLVYFNGRLAMSEDDLPYQVQITPNGDLKTVGRFDFDGOLESFMIAPKV 300  
Db 241 AHGTGVANAGLVYFNGRLAMSEDDLPYQVQITPNGDLKTVGRFDFDGOLESFMIAPKV 300  
Qy 301 DPESGELFALSYDVVSKPYLKYFRFSPDGTSPDVEIQLDQPTMMHDFAITENFVVVPDQ 360  
Db 301 DPESGELFALSYDVVSKPYLKYFRFSPDGTSPDVEIQLDQPTMMHDFAITENFVVVPDQ 360  
Qy 361 QVVFKLPEMIRGGSPVYVYDKNKVARFGILDKYADSSNIKWIDAPDCCFCHLNNAAEPE 420  
Db 361 QVVFKLPEMIRGGSPVYVYDKNKVARFGILDKYADSSNIKWIDAPDCCFCHLNNAAEPE 420  
Qy 421 TDEVVIGSCMTPTDPSIFNESDENLKSILSEIRLNLKTGSTRRPIISNEDQVNLBAGM 480  
Db 421 TDEVVIGSCMTPTDPSIFNESDENLKSILSEIRLNLKTGSTRRPIISNEDQVNLBAGM 480  
Qy 481 VNRNMLGRKTFAYLALAEAPKPVSGFAKVDLTGTVKHLVYGNRYGGSPFLPGBGGE 540  
Db 481 VNRNMLGRKTFAYLALAEAPKPVSGFAKVDLTGTVKHLVYGNRYGGSPFLPGBGGE 540  
Qy 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSRVYPYGHGTFIAGDDLAKQV 599  
Db 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSRVYPYGHGTFIAGDDLAKQV 599

## RESULT 2

US-09-758-269-16

; Sequence 16, Application US/09758269

; Patent No. US20020104120A1

; GENERAL INFORMATION:

; APPLICANT: IUCHI, SATOSHI

; APPLICANT: KOBAYASHI, MASATOMO

; APPLICANT: SHINOZAKI, KAZUO

; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN

; TITLE OF INVENTION: CLEAVAGE ENZYME GENE

; FILE REFERENCE: 3914-3

; CURRENT APPLICATION NUMBER: US/09/758,269

; CURRENT FILING DATE: 2001-01-12

; PRIOR APPLICATION NUMBER: JP 2001-003476

; PRIOR FILING DATE: 2001-01-11

; PRIOR APPLICATION NUMBER: JP 2000-010056

; PRIOR FILING DATE: 2000-01-13

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 16

; LENGTH: 605

; TYPE: PRT

; ORGANISM: Lycopersicon esculentum

US-09-758-269-16

Query Match 72.4%; Score 2280.5; DB 10; Length 605;  
Best Local Similarity 71.7%; Pred. No. 5.6e-221;  
Matches 435; Conservative 63; Mismatches 90; Indels 19; Gaps 6;

Qy 5 TATAVSGWLGNGHTQPPPLSSQSSDLSYCS---SLPWASRVTRKLVSSALHTPPALH 61  
Db 3 TTTSHATNTWI-----KTKLSMPFSKEFGFASNSISLLKNQNRQSLNINSLOAPPILH 57  
Qy 62 FPKQSN--SPA--IVVKKAKESN-----TKQMLFQRAAAALDAAEGELVSEKHLH 112  
Db 58 FPKQSNYQTPKNTWISHPKQENNNSSSTSKNLVQKAAWALDAVESALTKELEHP 117  
Qy 113 LPKTADPSVQIAGNFAPVNEQPVRRNLPVVGKLPDSIKGVYVRNGANLPHBPVTHHFD 172  
Db 118 LPKTADPRVQISGNFAPVNPENPVQSLPTGIPKVCQGVYVRNGANLPHBPVTHHFD 177

Qy 173 QDGMVHAVKPEHGSASYACRFTQTNRVQEROLGRPVFPFKKAIIGELHGTGIALMLFYAR 232  
Db 178 QDGMVHAVKPEHGSASYACRFTQTNRVQEROLGRPVFPFKKAIIGELHGTGIALMLFYAR 237  
Qy 233 AAAGTVDPAHGTGVANAGLVYFNGRLAMSEDDLPYQVQITPNGDLKTVGRFDFDGOLES 292  
Db 238 GLFGLVDHSGKTGVANAGLVYFNNRLAMSEDDLPYHVKVPTPTGDLKTGEGFDFDGOLES 297  
Qy 293 TMIAPKVPDPSGELFALSYDVVSKPYLKYFRFSPDGTSPDVEIQLDQPTMMHDFAIT 352  
Db 298 TMIAPKLPDPSGELFALSYDVVSKPYLKYFRFSPDGTSPDVEIQLDQPTMMHDFAIT 357  
Qy 353 NPVVVPDQVVKLPPEMIRGGSPVYVYDKNKVARFGILDKYADSSNIKWIDAPDCCFCHL 412  
Db 358 NPVVVPDQVVKLPPEMIRGGSPVYVYDKNKVARFGILDKYADSSNIKWIDAPDCCFCHL 417  
Qy 413 NNAAEETDETVVIGSCMTPTDPSIFNESDENLKSILSEIRLNLKTGSTRRPIISNEDQ 472  
Db 418 NNAAEETDETVVIGSCMTPTDPSIFNESDENLKSILSEIRLNLKTGSTRRPIISNEDQ 477  
Qy 473 QVNLBAGMVNRNMLGRKTFAYLALAEAPKPVSGFAKVDLTGTVKHLVYGNRYGGSPFL 532  
Db 478 QVNLBAGMVNRNMLGRKTFAYLALAEAPKPVSGFAKVDLTGTVKHLVYGNRYGGSPFL 537  
Qy 533 FLFPGS--GGEDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSRVYPYGHGTFI 590  
Db 538 FLFPGSKEEDDGYILAFVHDEKTKWSELQIVNAVSLVEATVKLPSRVYPYGHGTFI 597  
Qy 591 ADDLAKQ 597  
Db 598 ANDLANQ 604

## RESULT 3

US-09-758-269-12

; Sequence 12, Application US/09758269

; Patent No. US20020104120A1

; GENERAL INFORMATION:

; APPLICANT: IUCHI, SATOSHI

; APPLICANT: KOBAYASHI, MASATOMO

; APPLICANT: SHINOZAKI, KAZUO

; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN

; TITLE OF INVENTION: CLEAVAGE ENZYME GENE

; FILE REFERENCE: 3914-3

; CURRENT APPLICATION NUMBER: US/09/758,269

; CURRENT FILING DATE: 2001-01-12

; PRIOR APPLICATION NUMBER: JP 2001-003476

; PRIOR FILING DATE: 2001-01-11

; PRIOR APPLICATION NUMBER: JP 2000-010056

; PRIOR FILING DATE: 2000-01-13

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 12

; LENGTH: 612

; TYPE: PRT

; ORGANISM: Vigna unguiculata

US-09-758-269-12

Query Match 68.8%; Score 2168.5; DB 10; Length 612;  
Best Local Similarity 68.6%; Pred. No. 1.2e-209;  
Matches 423; Conservative 61; Mismatches 98; Indels 35; Gaps 8;

Qy 8 AAVSGWLGNGHTQPPPLSSQSSDLSYCS---SLPWASRVTRKLVSSALHTPPALHPPKQ- 65  
Db 3 SSASNTWFAVNLPSPPFKLPDSTS-SPTNLLPL--AKTSSNTTITCSLOT---LHPKQY 56  
Qy 66 ----SSNSPAIVVKKAKESNT-----KQNLFORAAAALDAA 100  
Db 57 QPTSTSTSTATTTTPTPIKTTTITTTTPPRETNPLSDTNQPLPQKWNFLQKAAATALDV 116  
Qy 101 EGLVSHKHLHLPKPTADPSVQIAGNFAPVNEQPVRRNLPVVGKLPDSIKGVYVRNGANP 160  
Db 117 ETALVSHKHLHLPKPTADPRVQIAGNFAPVPEHAADQGLPVVGKIPKCIDGVYVRNGANP 176

161	LHEV	TGHH	FDG	GMV	HA	VK	F	BH	G	S	A	S	A	C	R	T	Q	T	N	R	F	V	Q	B	R	Q	L	R	P	V	F	P	K	A	I	G	E	L	H	G	H	220														
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:															
177	LYE	V	A	G	H	H	F	D	G	M	V	H	A	V	K	F	T	Q	R	L	S	Q	B	K	S	L	G	R	P	V	F	P	K	A	I	G	E	L	H	G	H	236														
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:															
221	TG	I	A	R	L	M	F	Y	A	A	A	G	I	V	D	P	A	H	G	T	G	V	A	N	A	G	L	V	F	N	G	R	L	A	M	S	E	D	D	L	P	Y	O	V	Q	I	T	P	N	G	D	L	K	T	280	
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:				
237	S	G	I	A	R	L	L	F	Y	A	R	G	I	F	L	V	D	S	Q	M	G	V	A	N	A	G	L	V	F	N	N	H	L	A	M	S	E	D	D	L	P	Y	H	R	I	T	P	N	G	D	L	T	296			
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:				
291	V	G	R	F	D	F	G	O	L	E	S	T	M	I	A	H	P	K	V	D	P	S	G	E	L	F	A	L	S	I	D	V	S	K	P	L	K	P	R	F	S	D	G	T	K	S	P	D	V	E	I	Q	L	D	340	
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:				
297	V	G	R	D	F	N	G	L	N	S	T	M	I	A	H	P	K	L	D	P	V	D	G	L	H	A	L	S	I	D	V	I	O	K	P	L	K	P	R	F	S	D	G	V	K	S	P	D	V	E	I	P	L	K	356	
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:				
341	Q	P	T	M	H	H	F	A	I	T	E	N	F	V	V	P	D	Q	O	V	F	K	L	P	E	M	I	R	G	G	S	P	V	V	D	K	N	V	A	R	E	G	I	L	D	K	T	A	E	D	S	S	N	I	K	400
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:				
357	E	P	T	M	H	D	F	A	I	T	E	N	F	V	V	P	D	Q	O	V	F	K	L	P	E	M	I	R	G	G	S	P	V	V	D	K	N	T	S	R	F	I	L	H	K	N	A	K	A	D	A	N	A	M	R	416
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:				
401	W	I	D	A	P	D	C	F	P	L	N	N	A	E	E	P	E	T	D	E	V	V	I	G	S	C	M	T	P	D	S	I	F	N	E	S	D	E	N	L	K	S	V	E	I											

```

RESULT 4
US-09-758-269-2
; Sequence 2, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: TUCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patenlin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 583
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; US-09-758-269-2

```

117	SVVRNLTVBGTTPDCIDGVYIRNGANPMEPTAGHLLFDGDDGNVHAVKLTNGSASYACRF	176
194	TOTNREVOERQLGRPVFPFPAIXAGELHGHTGIARIMLFYARAAAGIVDPAHGTGVANAGLVY	253
177	TKTERLVQEKLGRRPVFPFPAIXAGELHGHTGIARIMLFYARGLCLGILNQNGVGVANAGLVY	236
254	FNGRLLAMSEDDLPYQVOITPNGDLKTVGRFPDQGLESTMTAHKPVKDPBESGELEFALSVD	313
237	FNNRLLAMSEDDLPYQLKITQGDUGTVGRYPDQGLKSGAMIAHPKLDPPVTKELHLSYD	296
314	VVSKPYLYKFRSPDCTGSPDVEIQDDQDTMMHDFAITENFVVVPQQVVFVKLPENIRGG	373
297	VVKPKYLYKFRSPDGVKSPELLEIPLETMTIHDFAITENFVVPDQVVVFVKLGEMISGK	356
374	SPVVYDNKVKARFGLDKYAEOSNSTKWIADPCFCFHLNAAWEEBETDEVVVIGSOWTP	433
357	SPVVFGEKYSRLGIMPKDXATEASQIIVNYSPTFCFHLNAAWESPETEEIVVIGSCMSP	416
434	PDSIFNESDENLKSVLSEIRLNLKTGESRTERPIISNEDQQVNLAEAGMWNRMILGRKTKFA	493
417	ADSIIFNRDEDSVLSIRINLURTKTRRSLLVNED--VNLIEGMVNRNRJGRKTRFA	474
494	YLALABPWPVKYSGFAKVDTLTGVBKXLYLDGNRYGGEPLFLPGEGG---EBDEGYILCF	549
475	FLAIAIWPVKYSGFAKVLDLCTGEMKKYVYGGEYGGEPFLPNSGNGBENEDDGYIFCH	534
550	VHDEKTKWSELQIWNAVSUEVATVKLPKSRVPYGFHGTFIGADDLAKQV	598
535	VHDEEKTSLQIINAVNLKLEATIKLPRVPYGFHGTTFVDSNELVDQL	583

```

RESULT 5
US-09-758-269-14
; Sequence 14, Application US/09758269
; Patent No. US20020104120A1
; GENERAL INFORMATION:
; APPLICANT: UCHI, SATOSHI
; APPLICANT: KOBAYASHI, MASATOMO
; APPLICANT: SHINOZAKI, KAZUO
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
; TITLE OF INVENTION: CLEAVAGE ENZYME GENE
; FILE REFERENCE: 3914-3
; CURRENT APPLICATION NUMBER: US/09/758,269
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: JP 2001-003476
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: JP 2000-010056
; PRIOR FILING DATE: 2000-01-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 14
; LENGTH: 604
; TYPE: PRT
; ORGANISM: Zea mays
US-09-758-269-14

```

Query Match	61.3%	Score	1930	DB	10	Length	604
Best Local Similarity	63.3%	Pred. No.	1.4e-185				
Matches	375	Conservative	73	Mismatches	122	Indels	22
Gaps	8						
QY	22	PPLSSSSDLSYCSGSLPMASRVRTKLNVSALHTP	--PALHFPFKOSSNSPAIW----	74			
DB	17	PARSARAS	-----NSVRFSPRAVSSVPPAECLQAPFKHPVADLPAPSRKFAAATVPQHA	71			
QY	75	-KPKAKESNTKQNNLFORAAAAALDA-AGEFLVS-HEKJLHPLPKTADPSVQIAGNFAPVN	131				
DB	72	AAPKAEGGKKQLNLFORAAAAALDAPEEGFVANVLERPHGLPSTADPAVOIAGNFAPVG	131				
QY	132	EQPVRENLPVVGKLPDSIKGVYVANGANELHEPTVTHHFGDGDGWHAVKEHGSA-SYA	190				
DB	132	ERPVPVHELPSVGRIPPTIDGVYARNGANPCEDPVAGHHILFGDGDGWHALRTRNGAASYA	191				
QY	191	CRFTQTNRVFOERQLGRFPVFPKATGELHIGHTGIALRLMLFYARAAAGIVDPAHGTVANAG	250				

192 CRFTETARLRQERAIGRFPFPAKAGELHGHSGIARLALFYARAAACGLVDPSAGTVANAG 251  
251 LAVFENGLLAMESEDDLPVQVOITENGDLKTGVRDFDQGLESTMIAPKPKVDPSGELFAL 310  
252 LVYFENGLLAMESEDDLPVHURVADDGDLTGVGRVDFDQGLGCAMIAHPKLDPATGELHAL 311  
311 SYDVUSKPYLYKFRPSDGTGKSPVEIQLDQPTMMHDFAITENFVVVDDQVQVFKPEMI 370  
312 SYDVIKRPLYKYFRPDGTGKSDVEIPLQOPTMIHDFAITENFVVVDDQVQVFKLQEML 371  
371 RGGSPVVVYDKNVAFGLDKYAESDNNIKWIDAPDCFCFHLNANAMEPEETDEVVVIGSC 430  
372 RGGSPVULDKERTGRFVLPKRAADASEMAWVDVDFCFHLNANAMEDEATGEVVVIGSC 431  
431 MTPDSIFNESDENLKSVLSEIRLNLTGSTRRPPIISNEDQOVNLEAGVMNMLGRKT 490  
432 MTPADSIFNESDERLESVLTEIRLDARTGRSTRRAVL P-PSQENLEVGMVNRNLLGRES 490  
491 KFAYLALAEAPKPKYSGFAKQDLTTGCEVKKHLVGNRYGGEPLFLPGEGBE-EDEGY 545  
491 RYAYLAVAEAPKPKESGFAKEDLSTGELTKFYEGGRFGGEPFVMPDPAAHPRGEDGY 550  
546 ILCFVHDEBKTWSELQIVNAVSLVEATVKLPSRVPYGFHGTFTGADDLAQV 597  
551 VLTFWHDERAGTSELLVYNAADIRLEATVQLPSRVPYGFHGTFTGQLEBAQ 602

## RESULT 6

US-09-758-269-10  
; Sequence 10, Application US/09758269  
; Patent No. US20020104120A1  
; GENERAL INFORMATION:  
; APPLICANT: IUCHI, SATOSHI  
; APPLICANT: KOBAYASHI, MASATOMO  
; APPLICANT: SHINOZAKI, KAZUO  
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN  
; FILE REFERENCE: 3914-3  
; CURRENT APPLICATION NUMBER: US/09/758,269  
; PRIOR FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: JP 2001-003476  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: JP 2000-010056  
; PRIOR FILING DATE: 2000-01-13  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 577  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-09-758-269-10

Query Match 52.8%; Score 1663.5; DB 10; Length 577;  
Best Local Similarity 55.4%; Pred. No. 1.1e-158;  
Matches 323; Conservative 100; Mismatches 137; Indels 23; Gaps 9;  
24 LSSQSDLSYCSSLPMASTRVTKLVSS-ALHTPALHPFKQSSNSPAIVVVKPKAKESN 82  
9 LLPTKTSRSHLLPQPKNANIGRRILINPFKIPTLPTDLSVPSP-----VLLKPTYPN 52  
83 TKOMILFORAAALDAAE-GFLVSHKELHPLPKTADPSVQIAGNFAPVNEQVRNLPV 141  
63 ---LNLQKLAATMLDKIESSIVPMEQNRPLPKPTDPAVLQSGNFAPVNECPVQNGLEV 119  
142 VGKLPDSIKGVYVRNANPLHEPVTHHFFDGGMWYAK--FEHGSASYACRTQTRNF 199  
120 VGOIPSLKGVYVRNANPNMFPPLAGHLHFDGDMTHAVS1GFDN-QVSSCRYTKTNRL 178  
200 VOERQLGRPVFPAKAGELHGHSGIARLMLFYARAAAGIVDPAHGTGVANAGLVYFNGRLL 259  
179 VQETALGRSVFPKAGELHGHSGIARLALFTABAGIGLVDTGRGMVANAGVVFNGRLL 238

260 AMSSEDDLPYQVOITENGDLKTGVRDFDQGLESTMIAPKPKVDPSGELFALSYDVUSKPY 319  
239 AMSSEDDLPYQKIDQGLTETIGRFGDDQSSVIAHPKPKVDATIGLHLSYVNLKPKH 298  
320 LKYFRPSDGTGKSPDVEIQLDQPTMMHDFAITENFVVVDDQVQVFKPEMIRGSGSPVYD 379  
299 LYLKFNCTCGKTRDVEITLPEPTMIHDFAITENFVWIPDQVQVFKLSEMRGSGSPVIYV 358  
380 KUKVARFGLDKYAESDNNIKWIDAPDCFCFHLNANAMEPEETDE-----VVVIGSCMTPD 435  
359 KEKMARFGLVSKQDLTGSDDINWVDVDFCFHLNANAMEE-RTBEGDPPVIVVIGSCMSPD 417  
436 STFNESDENLKSVLSEIRLNLTGSTRRPPIISNEDQOVNLEAGVMNMLGRKT KPAYL 495  
418 TIFSESGEPTRELSEIRLNLTGSTRRPPIISNEDQOVNLEAGVMNMLGRKT KPAYL 473  
496 ALAEAPKPKYSGFAKQDLTTGCEVKKHLVGNRYGGEPLFLPGEGBEEDGYLTCFVHDEKT 555  
474 AIADPWPCKSGIAKVDIONGTVSEFNYGSRFGGEPFVMPDPAAHPRGEDGYVNGVRDEEK 533  
556 WKSELQIVNAVSLVEATVKLPSRVPYGFHGTFTGADDLAQV 598  
534 DESEFVVVDATDMKQVAAVRLPERVPYGFHGTFTGQLEBAQ 576

## RESULT 7

US-09-758-269-4  
; Sequence 4, Application US/09758269  
; Patent No. US20020104120A1  
; GENERAL INFORMATION:  
; APPLICANT: IUCHI, SATOSHI  
; APPLICANT: KOBAYASHI, MASATOMO  
; APPLICANT: SHINOZAKI, KAZUO  
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN  
; FILE REFERENCE: 3914-3  
; CURRENT APPLICATION NUMBER: US/09/758,269  
; CURRENT FILING DATE: 2001-01-12  
; PRIOR FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: JP 2001-003476  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: JP 2000-010056  
; PRIOR FILING DATE: 2000-01-13  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 595  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-09-758-269-4

Query Match 30.7%; Score 968; DB 10; Length 595;  
Best Local Similarity 36.7%; Pred. No. 2.2e-88;  
Matches 210; Conservative 106; Mismatches 220; Indels 36; Gaps 13;  
42 SRVTRKLVSSALHTPALHPFKQSSNSPAIVVVKPKAKESNTKQMLFORAAALDAAE 101  
42 SPITNPSDNNDRNKPKTLH---NRTNHTLVSPPKLREEMTLATLAF---TTVEDVIN 94  
102 GFLVSHKELHPLPKTADPSVQIAGNFAPVNEQVRNLPV-GKLPDSIKGVYVRNANP 160  
95 TFDIP-----PSRPSVDPKHVLSDNPAVLDELPTDCEIHHGTLPLSLNAGVIRNPNP 149  
161 LHEPVTGHHFFDGGMWYAKVFEHGSASYACRTQTRNFVQERQLGRPVFPKAGELHGH 220  
150 QFLPRGPYHLFDGQGLHAIKIHNGKATLCSRVRVTKYKYNVEKQOTGAPVNPVFSGFNGV 209  
221 T-GTARLMLFYARAAAGIVDPAHGTGVANAGLVYFNGRLLAMSEDDLPYQVOITPKNGLK 279  
210 TASVARGALTAAVRLTGQYVNVNGIGLANTSLAFFNRLPALGESDLPFAVRLTESGDIE 269  
280 TVGRFDFDQGLESTMIAPKPKVDPSGELFALSYDVUSKPKYKFRPSDGTGKSPDVEI-Q 338  
270 TIGRYDFDQGLAWSMTAHPKTDITGBTFAFRYGPV-PPFLTYFRFDSAGKQORDVPIFS 328





Db 344 EDEVLITCRLENPDLDWMSGVKKELENFNGNELYEMFNMTKTSASQKLSASA-----LTTGEVKKHLY-- 522  
Qy 476 LEAGWNRNMLGRKTKFAYLALAEPPWPKVSGFAKVD-----LTTGEVKKHLY-- 522  
Db 399 VDFPRINECTYKQKQRYVGTILDSIAKVTGIKFDLHAEAFETGKRMLEVGNGIKGIYDL 458  
Qy 523 GDNRYGGEPLFLPGEGGEDEGYILCFVHDEKTKWKSBLQIVNAVSLVE--ATVKLPSRV 580  
Db 459 GEGRYGSEAIYVPRETAEDDGYLFFVHDENTGKSCVTVIDAKTMSAEPVAVVELPHRV 518  
Qy 581 PYGPHGTFIGADDLAKQV 599  
Db 519 PYGPHALFVTEEQLEQTL 537

## RESULT 10

US-09-758-269-33  
; Sequence 33, Application US/09758269  
; Patent No. US20020104120A1  
; GENERAL INFORMATION:  
; APPLICANT: IUCHI, SATOSHI  
; APPLICANT: KOBAYASHI, MASATOMO  
; APPLICANT: SHINOZAKI, KAZUO  
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN  
; FILE REFERENCE: 3914-3  
; CURRENT APPLICATION NUMBER: US/09/758,269  
; PRIOR FILING DATE: 2001-01-12  
; PRIOR APPLICATION NUMBER: JP 2001-003476  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: JP 2000-010056  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 33  
; LENGTH: 538  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-09-758-269-33

Query Match 29.7%; Score 936; DB 10; Length 538;  
Best Local Similarity 37.0%; Pred. No. 3.2e-85;  
Matches 207; Conservative 105; Mismatches 199; Indels 48; Gaps 13;  
Qy 64 KQSSNSPAIVVKPK-AKESNTKQMLFORAAAAALDAEGLVSHKHLPLPKTADPSVQ 122  
Db 4 KLSDCGIIISVHPRPSKGFSSKLLDLRLVVKLM-----HDSALPLH-----Y 47  
Qy 123 IAGNFAVP-NEOPVRNLPVVGKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMHVK 181  
Db 48 LSGNFAPIRDETFFPKDLPVHGFLPECLNGEFVRVGNPKFDVAGVYHWFDDGGMIRGVR 107  
Qy 182 FEHGSASACRTQTNRFVQERQLGRFPFPAKGAIGELHGTGIARLMLFYARAAAGIVDPA 241  
Db 108 IKDGKATYVSRYVTKSRKQEEFFCAKFMK-IGDLKGFGLLVVQQLRKLKILDNT 166  
Qy 242 HGTGVANAGLVFNGLLMSBDDLPYQVQITPNGLDKTVGRFPDGOLESTMAHPKV 301  
Db 167 YNGTANTALVYHGGKLLALQADKPYVVKLEDDGLQTLGIYDKELTHSFTAHPKVD 226  
Qy 302 PSBSGLFALSVDWSKPYLYKFRFPDGTGKSPDVEIQDQTMHDFAITENFVVPDQ 361  
Db 227 PYTGEMFTFGYS-HTPPYLYTYRIVSKDGMHDPVPIITSEPMHMDFAITETAYFMDLP 285  
Qy 362 VYFKLPEMIRGGSVV-YDNKVARFGILDKVAEDSSNIKVIDAPDCFCFLHNAWESE 420  
Db 286 MHFRPKRMVKEKMTYSDFTKARFGVLPYAKDELIRWFELENCPIFFNANAW--E 343  
Qy 421 TBEVVVIGSCMTPD-----SIFNESDENLKVLSIRNLKGTSTRRPIISNEDQVN 475  
Db 344 EDEVLITCRLENPDLDWMSGVKKELENFNGNELYEMFNMTKTSASQKLSASA----- 398

Qy 476 LEAGWNRNMLGRKTKFAYLALAEPPWPKVSGFAKVD-----LTTGEVKKHLY-- 522  
Db 399 VDFPRINECTYKQKQRYVGTILDSIAKVTGIKFDLHAEAFETGKRMLEVGNGIKGIYDL 458  
Qy 523 GDNRYGGEPLFLPGEGGEDEGYILCFVHDEKTKWKSBLQIVNAVSLVE--ATVKLPSRV 580  
Db 459 GEGRYGSEAIYVPRETAEDDGYLFFVHDENTGKSCVTVIDAKTMSAEPVAVVELPHRV 518  
Qy 581 PYGPHGTFIGADDLAKQV 599  
Db 519 PYGPHALFVTEEQLEQTL 537

## RESULT 11

US-08-976-063C-22  
; Sequence 22, Application US/08976063C  
; Publication No. US20020182697A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexander Steinbuechel; Horst Priefert; Jurgen Rabenhorst  
; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF  
; TITLE OF INVENTION: CONIFERYL ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN  
; TITLE OF INVENTION: ACID AND THEIR USE  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SPRUNG KRAMER SCHAEFER & BRISCOE  
; STREET: 660 White Plains Road  
; CITY: Tarrytown  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10591-5144  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage  
; COMPUTER: HP VECTRA  
; OPERATING SYSTEM: DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/976,063C  
; FILING DATE: 21-NOV-1997  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 196 49 655.1 (Germany)  
; FILING DATE: 29-NOV-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kurt G. Briscoe  
; REGISTRATION NUMBER: 33,141  
; REFERENCE/DOCKET NUMBER: Bayer 9998-CAO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (914) 332-1700  
; TELEFAX: (914) 332-1844  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 505 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-976-063C-22

Query Match 13.1%; Score 411.5; DB 8; Length 505;  
Best Local Similarity 25.8%; Pred. No. 2.8e-32;  
Matches 137; Conservative 79; Mismatches 192; Indels 123; Gaps 19;

Qy 118 DRSVQIAGNFAVPNEQVRNLPVVGKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMV 177  
Db 8 DP--QLVGTLLPRIEADLFDLEVDGEI PKSIGNIFYRNTPEPTPQKPFHTFIDGDGMA 65  
Qy 178 HAVKFEHGSASACRTQTNRFVQERQLGRFPFPAKGAIGELHGTGIARLMLFYARAAAGI 237  
Db 66 SAPHFEDGHVDFISRWVKTARFABELARKSLFGMYRNPYTDTSVKGL----- 114  
Qy 238 VDPAGHTGVANAGLVFNGLLMSBDDLPYQVQITPNGLDKTVGRFPDGOLES-TWIA 296  
Db 115 -----DRTVANTSIISHHGKVLAVKEDGLPY--ELDP-PLTRGHFDYDGGVTSQHTA 166



APPLICANT: BACHMANN, Heinrich  
APPLICANT: BRUGGER, Roland  
APPLICANT: FRIEDLEIN, Arno M  
APPLICANT: WIRTZ, Gabriele M  
APPLICANT: WOGGON, Wolf-Dietrich  
APPLICANT: WYSS, Adrian  
APPLICANT: WYSS, Markus  
TITLE OF INVENTION: BETA, BETA-CAROTENE 15, 15'-DIOXYGENASES, NUCLEIC ACID  
TITLE OF INVENTION: SEQUENCES CODING THEREFOR AND THEIR USE  
FILE REFERENCE: B-B-CAROTENE 15, 15'-DIOXYGENASES, ...  
CURRENT FILING DATE: 2002-01-15  
PRIOR APPLICATION NUMBER: 103382.0  
PRIOR FILING DATE: 1999-02-22  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 4  
LENGTH: 506  
TYPE: PRT  
ORGANISM: CHICKEN  
US-10-053-192-4

Query Match 7.8%; Score 247; DB 15; Length 506;  
Best Local Similarity 24.3%; Pred. No. 1.28-15;  
Matches 130; Conservative 83; Mismatches 213; Indels 110; Gaps 29;  
QY 132 EQPVERNLVVGKLPDSIKGVVYRNGANPLHE--PVTGHHFFDGDGMVHAKPEHGSASY 189  
Db 2 EHPFPIKAEOGQLTWLGVLRLNPG-MHTIGDTKYNHWFGLLHSHFTPKNGEVY 60  
QY 190 ACRF-----TOTNFEVERQLGRVFPKAGELHGHGTGIALMLFYARAAAGIVDP 240  
Db 61 RSKYLRSDTYNCIBANRIVVS-BEGTWAYDPCKNI-----FAKAFSYLSHTIPEFTDN 114  
QY 241 AHGTGVANAGLVYNGRLAMSEDDLPYQVITNGDLKTGVRFPDQGLESTW-IAHPK 299  
Db 115 CL-INIMKTGDY-----ANSETN--FIRKIDPQ-TLETLDKYDYKYAVANLATSPH 165  
QY 300 VDPESGELFALSYDVVSKYKIFRFS-----PDGTSKD-----VEI-----QLDQPT 343  
Db 166 YD-SAGNINMGTSIVDXGRTKYVLFKIPSSVPEKEKKSCFKHLEVVCSPISPSRLQPS 224  
QY 344 MMHDFAITENFVVPDQVWFKL-----PEMIRG-----GSPVYDKNVARFGILDYAE 394  
Db 225 YHSGFGLTENYIVFTEQ--FKLDIVKLATAYIRGVNWNASCLSFHKEDKTFHFVDRKTK 282  
QY 395 DSSNIX-WIDAPDCFCFLHNAWEEPTDEVVIGSCWTPPDSPN-----ESD 442  
Db 283 KEVSTKFTYDA--LVLYHHINAYEE--DGHVVPDIVAYRDNLSYDMFYLLKKLDKDEVN 337  
QY 443 ENLKSVL-----SEIRNL-----KTGSTRPPIISNEDQOVNL 476  
Db 338 NKLTSIPTCKRFVPLQVDKDEVGNLVKLETSATAVEKDGSIYCOPEILCE-----GI 393  
QY 477 EAGMVRNMLGRKTKFAYLALAEPPKVSFGFAKVDLTGTEVKKHLYGDNRYGGEPFLFPG 536  
Db 394 ELPRVNYDYGKKYKYVATEVQWSPVPTKIAKLVQTKV--LHWGEDHCWSPSEPIFVPS 452  
QY 537 -EGGDEGEYIL-CFVHDEKTKWSELOIVNAVSLV--EATVKLPSRVYPYGHGTF 588  
Db 453 PDAREDEGVWLTVCVVVSEPNKAPFLILLDAKTFKELGRATVNVENHL--DLHGMF 506

RESULT 15  
US-10-168-517-19  
Sequence 19, Application US/10168517  
Publication No. US20030166595A1  
GENERAL INFORMATION:  
APPLICANT: Von Lintig, J.  
APPLICANT: Vogt, K.  
TITLE OF INVENTION: No. US20030166595A1el dioxygenases catalyzing cleavage of beta-ca  
FILE REFERENCE: S-60094USPCT  
CURRENT APPLICATION NUMBER: US/10/168,517

CURRENT FILING DATE: 2002-06-19  
PRIOR APPLICATION NUMBER: 00105822.1  
PRIOR FILING DATE: 2000-03-20  
PRIOR APPLICATION NUMBER: 99125895.5  
PRIOR FILING DATE: 1999-12-24  
NUMBER OF SEQ ID NOS: 27  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 19  
LENGTH: 549  
TYPE: PRT  
ORGANISM: Danio rerio  
US-10-168-517-19  
Query Match 7.6%; Score 238.5; DB 12; Length 549;  
Best Local Similarity 21.7%; Pred. No. 9.7e-15;  
Matches 118; Conservative 89; Mismatches 219; Indels 119; Gaps 25;  
QY 130 VNEQPVRLNLPVVGKLPDSIKGVVYRNGANPLHEPVTCH-HFFDGDGMVHAKPEHGSAS 188  
Db 34 VEEIPDPITTLIKQIPSWINGSFLRNGPGKFEFGESKFTHWFGMALMHRFNKDGQVT 93  
QY 189 YACRFTQTRNFVQERQLGRVPPKKAIGELHGHGTGIALMLFYARAAAGIVDPAHGTGVAN 248  
Db 94 YSRRLQSDSYVQNSEKNRIVVSE-FGTLATPDPCNI---PARFERSFOIP-KTTDNAG 148  
QY 249 AGLVTFNGELLAMSEDDLPYQVITNGDLKTGVRFPDQGLE-STMIAHPKVDPESEL 307  
Db 149 VNFVYKSGDYVYSTETN--FMKIDP-VSLETKEKVDWSKFIASAAATAHPHYDRE-GAT 204  
QY 308 FALSYDVVSKYKLPYFRFSPDGTSKD-----VEIQLD-----OPTMHEDFAITENF 354  
Db 205 YMGNSYGRKGFYHILRVPPGKQDDADLSGAELCSIPAADPRKPSYVHSFVMSY 264  
QY 355 VVVPDQVWFKLPEM-----IRGGS---PVVDKNKVARFGILDKYAESSNIKWIDAPDC 407  
Db 265 IVFIEQPIKLDLTKFMYLXIAGSKFKVMSWNPEDTIFHVADRHTGQLLNTKIYSSA-M 323  
QY 408 FCFHLMNAWEEPTDEVVIGSCWTPPDSPN-----SDENLKSIVSE-----RVGG 529  
Db 324 FALHQNAYEE---NGYLIMDCCGDDGNVIGFLENLQSTGEDLDKFFNSLCTNLP 380  
QY 452 --IRNLKTGSTRPPIIS-----NED--QOVNLEAGMVRN-R 483  
Db 381 VYLPLEVEKDEPNQNLINLPTTASAVKTQTVGFVLYHEDLYNDDLLQYGGLEFPQINYA 440  
QY 484 NMLGRKTKFAYLALAEPPKVSFGFAKVDLTGTEVKKHLYGDN-----HVF 529  
Db 441 NYNARYPIFY-----ACGFG-----HVF 480  
QY 530 ---EPLFLPG-EGGDEGEYILCFVHDEKTKWSE-LOIVNAVSLVLEATVKLPSRVYPY 583  
Db 481 LPSPSEVFIAPDAQDEDDGVVMSVITPREKKSFPLLVLDKATFTELGRABVPVDIPYG 540  
QY 584 FHGTF 588  
Db 541 THGLF 545

Search completed: November 14, 2003, 21:23:16  
Job time : 38 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 21:15:22 ; Search time 21 Seconds  
(without alignments)  
1206.867 Million cell updates/sec

Title: US-09-758-269-6  
Perfect score: 3150  
Sequence: 1 MASFTATAVSGRWLGNHT.....VPYGHGTGIGADDLAKQVV 599

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCITUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	411.5	13.1	505	4	US-08-976-063E-22
2	210	6.7	533	3	US-08-385-259-2
3	210	6.7	533	4	US-08-645-370-2
4	209	6.6	533	1	US-08-488-305A-6
5	123.5	3.9	1043	3	US-08-928-361B-30
6	123.5	3.9	1721	3	US-08-700-651-5
7	123.5	3.9	1721	3	US-08-928-361B-6
8	123.5	3.9	1721	4	US-08-588-995A-6
9	116.5	3.7	1042	3	US-08-928-361B-11
10	116.5	3.7	1042	4	US-08-588-995A-11
11	116.5	3.7	1837	3	US-08-928-361B-5
12	116.5	3.7	1837	4	US-08-588-995A-5
13	112	3.6	60	4	US-08-311-731A-403
14	106	3.4	412	3	US-08-461-697-77
15	105.5	3.3	471	4	US-08-107-532A-5705
16	105	3.3	1118	4	US-08-585-173B-36
17	104.5	3.3	462	4	US-08-328-352-7128
18	104.5	3.3	2172	1	US-08-611-107-31
19	102	3.2	1131	4	US-08-252-991A-31247
20	101.5	3.2	3340	4	US-08-252-991A-23568
21	98	3.1	741	3	US-08-001-984C-106
22	98	3.1	741	4	US-08-396-347F-106
23	97.5	3.1	1031	4	US-08-585-173B-40
24	97.5	3.1	10182	4	US-08-134-001C-3159
25	97	3.1	516	1	US-08-097-829-4
26	97	3.1	516	1	US-08-577-403-4
27	96	3.0	213	4	US-08-252-991A-28509

28 96 3.0 1410 3 US-09-335-409-3 Sequence 3, Appli  
29 96 3.0 1410 4 US-09-568-102-3 Sequence 3, Appli  
30 96 3.0 1410 4 US-09-567-969-3 Sequence 3, Appli  
31 96 3.0 1410 4 US-09-568-480-3 Sequence 3, Appli  
32 96 3.0 1410 4 US-09-568-486-3 Sequence 3, Appli  
33 96 3.0 1410 4 US-09-568-472-3 Sequence 3, Appli  
34 96 3.0 1410 4 US-09-567-899-3 Sequence 3, Appli  
35 95.5 3.0 1802 4 US-09-322-478-18 Sequence 18, Appli  
36 94.5 3.0 620 3 US-09-000-145-5 Sequence 5, Appli  
37 94.5 3.0 1297 4 US-09-107-532A-4552 Sequence 4552, Ap  
38 94 3.0 1471 3 US-08-755-587-188 Sequence 188, App  
39 92.5 2.9 868 2 US-08-907-166-2 Sequence 2, Appli  
40 92.5 2.9 868 4 US-09-391-340-2 Sequence 2, Appli  
41 92.5 2.9 882 4 US-09-328-352-5232 Sequence 5232, Ap  
42 92.5 2.9 1057 4 US-09-697-367-23 Sequence 23, Appli  
43 92.5 2.9 1057 4 US-09-394-272-2 Sequence 2, Appli  
44 92 2.9 205 3 US-09-385-259-3 Sequence 3, Appli  
45 92 2.9 205 4 US-09-645-370-3 Sequence 3, Appli

#### ALIGNMENTS

##### RESULT 1

US-08-976-063E-22

; Sequence 22, Application US/08976063E

; Patent No. 6524831

; GENERAL INFORMATION:

; APPLICANT: Steinbuechel, Alexander

; APPLICANT: Priefert, Horst

; APPLICANT: Rabenhorst, Jurgen

; TITLE OF INVENTION: SYNTHETIC ENZYMES FOR THE PRODUCTION OF CONIFERYL

; TITLE OF INVENTION: ALCOHOL, CONIFERYLALDEHYDE, FERULIC ACID, VANILLIN AND

; FILE REFERENCE: VANILLIC ACID AND THEIR USE

; FILE REFERENCE: Bayer-9938-CAO

; CURRENT APPLICATION NUMBER: US/08/976,063E

; CURRENT FILING DATE: 1997-11-21

; PRIOR APPLICATION NUMBER: 196 49 655.1 GERMANY

; PRIOR FILING DATE: 1996-11-29

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 22

; LENGTH: 505

; TYPE: PRT

; ORGANISM: not required under old rule

US-08-976-063E-22

Query Match 13.1%; Score 411.5; DB 4; Length 505;

Best Local Similarity 25.8%; Pred No. 4,6e-34;

Matches 137; Conservative 79; Mismatches 192; Indels 123; Gaps 19;

QY 118 DFSVQIAGNFAPVNEQVVRNLPVVGKLPDSIKGVYVANGANPLHPTVGHFFPDGGMV 177

Db 8 DP-QLVGTLLPTR:EAOLFLEVDGEIPKSGINGTFYRNTPEPQVTPQKFTTFIDGDM 65

QY 178 HAVKEHSGASACACFTQTNRFVQERQLGRVFPKAIHELGHGTGIAIARLMLFYARAAGI 237

Db 66 SAHFEDGHVDVDSISRWKTAFTARLARKSLFGMYRNPYDDTSVKGL----- 114

QY 238 VDPFHGTGVANAGLYFNGRLIAMEDDLPYQVQITPNDGLKTGRFPDQGLS-TWIA 296

Db 115 -----DRTVANTSIIISHEGKVLAVKEDGLPY--ELDPK-TLETGRGHFDYDGVTSQHTA 166

QY 297 HKVPDPESGELFALSVDVVKPYLKYFRFSPDGTKSPDV-----EIQLDOP- 342

Db 167 HPKYDPETGDL-----FFGSAKGEATPMAYIYVDNKGKVTHTWTFEPY 213

QY 343 -TMMHDFATTENFVVPPDQVVFVKLPPEMIRGSPV-VYDNKNKVARFGILDYAESSNIK 400

Db 214 GAFMDFALTNRWS:FPIMPATNSL-SRLKAKQPIYMWPELGSYIGVL--APRQGSJLR 270

QY 401 WTDADPCFCFHUNAWEEPETDEVVVVIGSCWTP---PDS-----I 437

Db 401 WTDADPCFCFHUNAWEEPETDEVVVVIGSCWTP---PDS-----I 437



Db 442 FVPDRCKLVKTKETWVQEPDSYPSPIFVSHPDALBEDDGVLSVVVSPGAGQKPAY 501  
QY 560 LQIVNAVSLEVEATVKLPSRVPYGFHGTFF 588  
Db 502 LLILNAKDLSEVARAEVEINIPVTFHGLF 530

RESULT 4  
US-08-488-305A-6  
; Sequence 6, Application US/08488305A  
; Patent No. 5679772  
; GENERAL INFORMATION:  
; APPLICANT: B vik, Claes Olof, Eriksson, Ulf; Peterson, Per A.  
; TITLE OF INVENTION: Isolated Protein Receptors, Antibodies Which  
; TITLE OF INVENTION: bind Thereto, Nucleic Acid Sequence Coding  
; Patent No. 5679772  
; TITLE OF INVENTION: Therefor, And Uses Thereof  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Felfe & Lynch  
; STREET: 805 Third Avenue  
; CITY: New York City  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 kb storage  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/488,305A  
; FILING DATE: 7-JUNE-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kohli, Vineet  
; REGISTRATION NUMBER: 37,003  
; REFERENCE/DOCKET NUMBER: LUD 5280.3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 688-9200  
; TELEFAX: (212) 838-3884  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 533 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-488-305A-6

Query Match 6.6%; Score 209; DB 1; Length 533;  
Best Local Similarity 20.7%; Pred. No. 8.7e-13;  
Matches 119; Conservative 99; Mismatches 222; Indels 134; Gaps 29;

QY 105 VSHEKLHPLKPTADPSVQIAGNAPVNEQPVRRNLP-----VVGKLPDSIKGVVVRNGANP 160  
Db 1 MSSQVEHP-----AGYKCLFETVEELSSPLTAHVTRGIPLWLTGSLRCGPGL 49  
QY 161 LH---EPVTGHHFDGGMHVAHFEGSASACRFTQTRNFQVE-----RQLGRPV 209  
Db 50 FEVGSERP--YHLDGQALLHKFPDKGHVYHRRFRTDAYVRAVTEKRVITEFTCA 107  
QY 210 PPKAIGELHGTGTLARMLFYARAAAGIVDPAHGTGVANAGLVYFNGRLLAMSED----- 264  
Db 108 FPDPCNIT-----FSRFFSYF-----RGVEVTDNALV-----NIYFVGEDYYACT 147  
QY 265 DLPYQVQITPNGDLTKVGRFDFDQGLE-STWIAHPKVDPS-----GELFALSVDV 314  
Db 148 ETNFIKKNPE-TLETIKQVLDLNVVSVNGATAPHPHENDGTVYNGNCFGKFSIAYNI 206  
QY 315 VSKPYLYKVRFPDGTSPDVEIQLD-----OPTMHDFAITENFVVVDPQVVKLPDEM 369  
Db 207 VKIPPLQADKEDP--ISKSEIWWQPCSDRFPKPSVHSFGLTPNYIVFVETPVKINLFPK 264  
QY 370 IRGGSPPVYDKNKVAF-----GILDKYAEADSSNKNKIDAPDCFCFHLNVAWEPP 419

Db 265 LSSWS--LWGANYVDCFESNETGVLHIAADKKKKYNNKYRTP--NFLFHHINTYEDH 321  
QY 420 ETDEVVIGSCMTDPDSIFNED--ENLKSVLSEIRLNLKGTGSTRPPIISNEDQVNL EA 478  
Db 322 EF--LIVDLCCWKGFVYNYLYLANLRNWEVEK--KNARKAPOPEVRYVLPNLNDK 376  
QY 479 GWNRRM--LGRKTKFAVLAALAE--W--PKV--SG-----FAKVDTLT--GEVKKHLYG 523  
Db 377 ADTGKLVTLPTTATILCSDETIWLEPEVLFSGPQAQEFPPQINVKYGGKPYVAYG 436  
QY 524 -----DNRYGGEPLFLP--GEGGDEBYILCFVHDEKTM 556  
Db 437 LGLNHFVDRCLKLVKTKETWVQEPDSYPSPIFVSHPDALBEDDGVLSVVVSPGAG 496  
QY 557 K--SLQIVNAVSLEVEATVKLPSRVPYGFHGTFF 588  
Db 497 QKPAYLLILNAKDLSEVARAEVEINIPVTFHGLF 530

RESULT 5  
US-08-928-361B-30  
; Sequence 30, Application US/08928361B  
; Patent No. 6071518  
; GENERAL INFORMATION:  
; APPLICANT: Petersen, Carolyn  
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,  
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS  
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM  
; TITLE OF INVENTION: SPECIES INFECTIONS  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA  
; STREET: 385 Sherman Avenue, Suite 6  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306-1840  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/928.361B  
; FILING DATE: 12-SEP-1997  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/026,062  
; FILING DATE: 13-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Verny, Hana  
; REGISTRATION NUMBER: 30,518  
; REFERENCE/DOCKET NUMBER: 480.76-1 (HV)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-324-1677  
; TELEFAX: 650-324-1678  
; INFORMATION FOR SEQ ID NO: 30:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1043 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-928-361B-30

Query Match 3.9%; Score 123.5; DB 3; Length 1043;  
Best Local Similarity 20.6%; Pred. No. 0.0026;  
Matches 125; Conservative 85; Mismatches 194; Indels 203; Gaps 35;

QY 5 TATAVSGRWLGGNHTQPPGLSSQSSDLSYCSSL--PMASRVTRK----LNVSSALHTPP 58  
Db 153 TIAGIVSG-----ISASESL-LSQKSALIDPATNMVVGEGGLLNPATGVMIFG 200





## ATTORNEY/AGENT INFORMATION:

NAME: Vetry, Hana  
 REGISTRATION NUMBER: 30,518  
 REFERENCE/DOCKET NUMBER: 480.76-1 (HV)  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 650-324-1677  
 TELEFAX: 650-324-1678  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1721 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-928-361B-6

Query Match 3.9%; Score 123.5; DB 3; Length 1721;

Best Local Similarity 20.6%; Pred. No. 0.0064;  
 Matches 125; Conservative 85; Mismatches 194; Indels 203; Gaps 35;

```

QY 5 TATAVSGRWLGNGNHTQPPSSQSSDLSYCSSL--PMASRVTRK-----LNVSSALHTPP 58
DB 831 TIAGIVSG-----ISASESL-LSQKSALIDPATNMVVGEGFGLLNPAATGVMI 878
QY 59 ALHFPKQSSNSPAI-----VVKPKAKESNTKQNNL-----FQRAAAALD 98
DB 879 FLGPEQTFQSFIEDGGIIPPEVAANADKFKLSIPSPVESIPEKDQKIDSISELMYD 938
QY 99 AAEGLVSHKHLPLKPTADPSVQIAGNFAPVNEQPVNRNLPVVGKLPDSIKGVYVRNGA 158
DB 939 IESGRLLIGQVSKRPIPGS-----IAGDLNPIKMTPTQTD-SVTGKPIDPTTGL 985
QY 159 NPLHEPVTGHHFFDGDGMVHAKVPEHGSASY--ACRFTQTNRFVQERQVGRPVFPKAIGE 216
DB 986 -PFNPP-TGH-----LINPTNNMTDSSPAGAYKAVSNGIKTDNVYGLPV-----GE 1031
QY 217 LHGHTGIARLMFLYARAAAAGIVDPAHGTGVAN--AGLVYFNGRL-LAMSEDDLPLYQVQIT 273
DB 1032 ITGLPKDPSDIPFNSTTIGELVDPSTGKPINNSTAGIV--SGKPGLPPIEDE-----1081
QY 274 PNGDLKTVGRFD-----FDGQLESTMTAHKPKVDPESELFALSVDYVSKPKYLKFRFSP 327
DB 1082 -NGNL-----FDPSTNLPIDGN-----NQLVNPETNSTVSGSTSGTKP-----KP 1121
QY 328 -----DGTKSPDVEI--QLDQ-----PTMMDHFAITENFVVVPDQVVF 364
DB 1122 GIPVNGGVVPDEAKQADKQKGLIVPPTNSINKDPVTNTQYNTTGNIIINPETGKV-1180
QY 365 KLPFEMIRG-----GSPVYVDKVKARFGILDKYAEDSSNIKWIDAPDCFCFLMNAWE 417
DB 1181 -IPGSLPGLSNLVPFNTPOQTDE-----ITGKPVDTVTGLPY-----1216
QY 418 EPETDEVVVGSCMTPPDSIFNESDENLKSVLSEIRNLKLTGSTRPPIISNEDQOVNLE 477
DB 1217 DPSTGEIIDPATKLPVPSVAGD-----EILTEV-LNITDEVTVGLPI-----DLE 1261
QY 478 AGVARNMLGRKTKFAYLALAEWPVKVSGFAK---DLTTGEVKKHLYGDNRYGGEPLFL 534
DB 1262 TGL-----PRDPVSGLPQLNGTLVDPNSKKPIPGS--HSG--FI 1297
QY 535 PEGGEE 541
DB 1298 NGTSGEQ 1304

```

## RESULT 8

US-09-588-995A-6  
 Sequence 6, Application US/09588995A  
 Patent No. 6514697  
 GENERAL INFORMATION:  
 APPLICANT: PETERSEN, CAROLYN  
 APPLICANT: BARNES, DEBRA A.  
 APPLICANT: NELSON, RICHARD C.

APPLICANT: GUT, JIRI  
 TITLE OF INVENTION: METHODS FOR DETECTION OF CRYPTOSPORIDIUM SPECIES AND  
 TITLE OF INVENTION: ISOLATES AND FOR DIAGNOSIS OF CRYPTOSPORIDIUM  
 TITLE OF INVENTION: INFECTIONS

FILE REFERENCE: 480.19-5  
 CURRENT APPLICATION NUMBER: US/09/588.995A  
 CURRENT FILING DATE: 2000-06-06  
 PRIOR APPLICATION NUMBER: 08/827,171  
 PRIOR FILING DATE: 1997-03-27  
 PRIOR APPLICATION NUMBER: 08/928,361  
 PRIOR FILING DATE: 1997-09-12  
 PRIOR APPLICATION NUMBER: 08/700,651  
 PRIOR FILING DATE: 1996-08-14  
 PRIOR APPLICATION NUMBER: 08/415,751  
 PRIOR FILING DATE: 1995-04-03  
 NUMBER OF SEQ ID NOS: 115  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 6  
 TYPE: PRT  
 LENGTH: 1721  
 ORGANISM: Cryptosporidium parvum  
 US-09-588-995A-6

Query Match 3.9%; Score 123.5; DB 4; Length 1721;

Best Local Similarity 20.6%; Pred. No. 0.0064;  
 Matches 125; Conservative 85; Mismatches 194; Indels 203; Gaps 35;

```

QY 5 TATAVSGRWLGNGNHTQPPSSQSSDLSYCSSL--PMASRVTRK-----LNVSSALHTPP 58
DB 831 TIAGIVSG-----ISASESL-LSQKSALIDPATNMVVGEGFGLLNPAATGVMI 878
QY 59 ALHFPKQSSNSPAI-----VVKPKAKESNTKQNNL-----FQRAAAALD 98
DB 879 FLGPEQTFQSFIEDGGIIPPEVAANADKFKLSIPSPVESIPEKDQKIDSISELMYD 938
QY 99 AAEGLVSHKHLPLKPTADPSVQIAGNFAPVNEQPVNRNLPVVGKLPDSIKGVYVRNGA 158
DB 939 IESGRLLIGQVSKRPIPGS-----IAGDLNPIKMTPTQTD-SVTGKPIDPTTGL 985
QY 159 NPLHEPVTGHHFFDGDGMVHAKVPEHGSASY--ACRFTQTNRFVQERQVGRPVFPKAIGE 216
DB 986 -PFNPP-TGH-----LINPTNNMTDSSPAGAYKAVSNGIKTDNVYGLPV-----GE 1031
QY 217 LHGHTGIARLMFLYARAAAAGIVDPAHGTGVAN--AGLVYFNGRL-LAMSEDDLPLYQVQIT 273
DB 1032 ITGLPKDPSDIPFNSTTIGELVDPSTGKPINNSTAGIV--SGKPGLPPIEDE-----1081
QY 274 PNGDLKTVGRFD-----FDGQLESTMTAHKPKVDPESELFALSVDYVSKPKYLKFRFSP 327
DB 1082 -NGNL-----FDPSTNLPIDGN-----NQLVNPETNSTVSGSTSGTKP-----KP 1121
QY 328 -----DGTKSPDVEI--QLDQ-----PTMMDHFAITENFVVVPDQVVF 364
DB 1122 GIPVNGGVVPDEAKQADKQKGLIVPPTNSINKDPVTNTQYNTTGNIIINPETGKV-1180
QY 365 KLPFEMIRG-----GSPVYVDKVKARFGILDKYAEDSSNIKWIDAPDCFCFLMNAWE 417
DB 1181 -IPGSLPGLSNLVPFNTPOQTDE-----ITGKPVDTVTGLPY-----1216
QY 418 EPETDEVVVGSCMTPPDSIFNESDENLKSVLSEIRNLKLTGSTRPPIISNEDQOVNLE 477
DB 1217 DPSTGEIIDPATKLPVPSVAGD-----EILTEV-LNITDEVTVGLPI-----DLE 1261
QY 478 AGVARNMLGRKTKFAYLALAEWPVKVSGFAK---DLTTGEVKKHLYGDNRYGGEPLFL 534
DB 1262 TGL-----PRDPVSGLPQLNGTLVDPNSKKPIPGS--HSG--FI 1297
QY 535 PEGGEE 541
DB 1298 NGTSGEQ 1304

```

## RESULT 9

US-08-928-361B-11  
; Sequence 11, Application US/08928361B  
; Patent No. 6071518  
; GENERAL INFORMATION:  
; APPLICANT: Petersen, Carolyn  
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,  
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS  
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM  
; TITLE OF INVENTION: SPECIES INFECTIONS  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA  
; STREET: 385 Sherman Avenue, Suite 6  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306-1840  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/928,361B  
; FILING DATE: 12-SEP-1997  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/026,062  
; FILING DATE: 13-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: VERNY, Hana  
; REGISTRATION NUMBER: 30,518  
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-324-1677  
; TELEFAX: 650-324-1678  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1042 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-928-361B-11  
Query Match 3.7%; Score 116.5; DB 3; Length 1042;  
Best Local Similarity 19.9%; Pred. No. 0.014;  
Matches 113; Conservative 74; Mismatches 175; Indels 207; Gaps 30;  
QY 48 LNVSSALHTPPALHFPKQSSNGPAI---VVKPKAKESNTKQML----- 88  
DB 190 LNPATGVMTPGSLGPSEQTFPFEIEDGGIIPPEVAANADKFKLSIPPVSPEIPEKDQ 249  
QY 89 -FORAAAAALDAAGFLVSHKHLPLPKTADPSVQIAGNFAPVNEQPVVRNLPVVGKLPD 147  
DB 250 KIDSISELMYDIESGRLLIGQVSKRPFGS-----IAGDLNPIKMTPTQTD-SVTGKPID 302  
QY 148 SIKGVVVRGANPLHPEVTHGHFFDGDGMVHAKFEHGSASY--ACRFTQTNRFVQERQL 205  
DB 303 PTTGL-----PFNPP-TGH-----LINPTNNTMDSFAGAYKAVSNGIKTDNVY 347  
QY 206 GRPV-----FPKAGELHGHGTGIARLMLFYARAAAGIVDPAHGTGVAN-- 248  
DB 348 GLPVDIEITGLPKDPVSDIPFNSITGEL-----VDPSTGKPINNYT 387  
QY 249 AGLYVYNGRLLAMSEDDLQVQVQITPNGDL---KTVGRFDFDGOLESTMIHAKVPDPESG 305  
DB 388 AGIV-SGKRGLPPIEDE-----NGNLFDPSTKLPIDGNQNL-----VNPETN 428  
QY 306 ELFALSYDVVSKPYLYFRFSP-----DGTKSPDVEI--QLDQ-----P 342  
DB 429 STVSGTSGSTKF-----KPGIPVNGGVPEEAKQADKGDGLIVPPTNSINKDP 481

QY 343 TMMHDFALTENFVVVDQVVKLPLEMRG-----GSPVYDKNKVAREFGILDKYAE 395  
DB 482 VTNTQISNTGNIINPFTGKV--IPGSLPGSLNVPFNTPOCTDE-----ITGKPYDT 532  
QY 396 SSNIKWIDAPDCFCFHLNNAWEPEPDETVVVGSCMTPTPDSIFNESDENLKSIVSEIRLN 455  
DB 533 VTGLPY-----DPSTGIIIDPATKLPFGSVAGD-----EILTEV-LN 569  
QY 456 LATGSTRPPLISNEDQOVNLEAGVMNRNMLGRKTKFAYLALABWPVKVSGFAKV---DL 512  
DB 570 ITTDEVTGLPI-----DLETGL-----PRDPVSLGLPQLPNTGL 602  
QY 513 TTGEVKGHLGDNRYGGEPLFLPGEGEE 541  
DB 603 VDPsNKKPIPGS--HSG---FINGTSGEQ 626  
RESULT 10  
US-09-588-995A-11  
; Sequence 11, Application US/09588995A  
; Patent No. 6514697  
; GENERAL INFORMATION:  
; APPLICANT: PETERSEN, CAROLYN  
; APPLICANT: BARNES, DEBRA A.  
; APPLICANT: NELSON, RICHARD C.  
; APPLICANT: GUT, JIRI  
; TITLE OF INVENTION: METHODS FOR DETECTION OF CRYPTOSPORIDIUM SPECIES AND  
; TITLE OF INVENTION: ISOLATES AND FOR DIAGNOSIS OF CRYPTOSPORIDIUM  
; TITLE OF INVENTION: INFECTIONS  
; FILE REFERENCE: 480.19-5  
; CURRENT APPLICATION NUMBER: US/09/588,995A  
; CURRENT FILING DATE: 2000-06-06  
; PRIOR APPLICATION NUMBER: 08/827,171  
; PRIOR FILING DATE: 1997-03-27  
; PRIOR APPLICATION NUMBER: 08/928,361  
; PRIOR FILING DATE: 1997-09-12  
; PRIOR APPLICATION NUMBER: 08/700,651  
; PRIOR FILING DATE: 1996-08-14  
; PRIOR APPLICATION NUMBER: 08/415,751  
; PRIOR FILING DATE: 1995-04-03  
; NUMBER OF SEQ ID NOS: 115  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 11  
; LENGTH: 1042  
; TYPE: PRT  
; ORGANISM: Cryptosporidium parvum  
US-09-588-995A-11  
Query Match 3.7%; Score 116.5; DB 4; Length 1042;  
Best Local Similarity 19.9%; Pred. No. 0.014;  
Matches 113; Conservative 74; Mismatches 175; Indels 207; Gaps 30;  
QY 48 LNVSSALHTPPALHFPKQSSNGPAI---VVKPKAKESNTKQML----- 88  
DB 190 LNPATGVMTPGSLGPSEQTFPFEIEDGGIIPPEVAANADKFKLSIPPVSPEIPEKDQ 249  
QY 89 -FORAAAAALDAAGFLVSHKHLPLPKTADPSVQIAGNFAPVNEQPVVRNLPVVGKLPD 147  
DB 250 KIDSISELMYDIESGRLLIGQVSKRPFGS-----IAGDLNPIKMTPTQTD-SVTGKPID 302  
QY 148 SIKGVVVRGANPLHPEVTHGHFFDGDGMVHAKFEHGSASY--ACRFTQTNRFVQERQL 205  
DB 303 PTTGL-----PFNPP-TGH-----LINPTNNTMDSFAGAYKAVSNGIKTDNVY 347  
QY 206 GRPV-----FPKAGELHGHGTGIARLMLFYARAAAGIVDPAHGTGVAN-- 248  
DB 348 GLPVDIEITGLPKDPVSDIPFNSITGEL-----VDPSTGKPINNYT 387  
QY 249 AGLYVYNGRLLAMSEDDLQVQVQITPNGDL---KTVGRFDFDGOLESTMIHAKVPDPESG 305  
DB 388 AGIV-SGKRGLPPIEDE-----NGNLFDPSTKLPIDGNQNL-----VNPETN 428  
QY 306 ELFALSYDVVSKPYLYFRFSP-----DGTKSPDVEI--QLDQ-----P 342

Db 429 STVSGTSGSTKP-----KPGIPVNGGVVDEEAKQADKQDGLIVPPTNSINKDP 481  
QY 343 TMHEDFAITENFVVDPDQVVFVKLPEMIRG-----GSPVYVDKNKVARFGLDKYABD 395  
Db 482 VTNTQVSNNTGNIINPETGKV--IPGSLFGSLNYPSPNTPOOTDE-----ITGKPVDT 532  
QY 396 SSNIKIDAPDCFCFHLNABEPETDEVVIGSCMTPPDSIFNESDENLKSIVLSEIRLN 455  
Db 533 VTGLPY-----DPSTGEIIDPATKLPFGSVAGD-----EILTEV-LN 569  
QY 456 LKTGESSTRAPITISNEDQOVNLEAGVNRNMLGRKTKFAYLALAEWPVKVSGPAKV---DL 512  
Db 570 ITTDEVVTGLPI-----DLETGL-----PRDPVSGLPQLPNGTL 602  
QY 513 TTGEVKKHLYGDNRYGGEPLFLPGEEGEE 541  
Db 603 VDPSNKKPIFGS--HSG---FINGTSGEQ 626

## RESULT 11

US-08-928-361B-5  
; Sequence 5, Application US/08928361B  
; Patent No. 6071518  
; GENERAL INFORMATION:  
; APPLICANT: Petersen, Carolyn  
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,  
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS, AND FRAGMENTS  
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM  
; TITLE OF INVENTION: SPECIES INFECTIONS  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PETERS, CARYN, JONES & BIKSA  
; STREET: 385 Sherman Avenue, Suite 6  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306-1840  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/928,361B  
; FILING DATE: 12-SEP-1997  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/026,062  
; FILING DATE: 13-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Vervy, Hana  
; REGISTRATION NUMBER: 30,518  
; REFERENCE/DOCKET NUMBER: 480.76-1 (HV)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-324-1677  
; TELEFAX: 650-324-1678  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1837 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-928-361B-5

Query Match 3.7%; Score 116.5; DB 3; Length 1837;

Best Local Similarity 19.9%; Pred. No. 0.039;  
Matches 113; Conservative 74; Mismatches 175; Indels 207; Gaps 30;

QY 48 LNVSSALHTPPALHFFKQSNSPA1----VVKPKAKESNTKQVNL-----88  
Db 985 LNPATGWIPIGSLGPEQTPFSIEIDGGIIPPEVAANADKFKLSIPSPVESIPEDQ 1044

QY 89 -FQRAAAALDAABEGFLVSHKHLPLPKTADPDSVQIAGNAPVNEQVRANLPLVVGKLPD 147  
Db 1045 KIDSISBLMDIESGRLLIGQVSKPIFGS-----TAGDLNLPINKTKTQTD-SYTGKPID 1097  
QY 148 SIKGVYVRNGANPLHEHVTGTHHFPDGDGMVHAVKFEHGSASY--ACRFTOTNRREVQRQL 205  
Db 1098 PTTGL-----PNPP-TGH-----LINPNNNTMDSFAGAYKXVNSGKTKDNVY 1142  
QY 206 GRPV-----FPKAIHELHGTGIAARMLFYARAAAGLIVPAHGTGVAN-- 248  
Db 1143 GLPVDEITGLPKDPVSDIPFNSTTGL-----VDPSTGKPINNYT 1182  
QY 249 AGLVYFNGRLLAMSEDDLPYQVQITPNGDL---KTVGRFPDGDQLESTMIAPKVDPESG 305  
Db 1183 AGIV-SGKRGLPPIEDE-----NGNLFDSSTKLPIDGNQL-----VNPETN 1223  
QY 306 ELFALSVDVSKPYLYKFRFSP-----DGTKSDDVEI--QLDQ-----P 342  
Db 1224 STVSGTSGSTKP-----KPGIPVNGGVVDEEAKQADKQDGLIVPPTNSINKDP 1276  
QY 343 TMHEDFAITENFVVDPDQVVFVKLPEMIRG-----GSPVYVDKNKVARFGLDKYABD 395  
Db 1277 VTNTQVSNNTGNIINPETGKV--IPGSLFGSLNYPSPNTPOOTDE-----ITGKPVDT 1327  
QY 396 SSNIKIDAPDCFCFHLNABEPETDEVVIGSCMTPPDSIFNESDENLKSIVLSEIRLN 455  
Db 1328 VTGLPY-----DPSTGEIIDPATKLPFGSVAGD-----EILTEV-LN 1364  
QY 456 LKTGESSTRAPITISNEDQOVNLEAGVNRNMLGRKTKFAYLALAEWPVKVSGPAKV---DL 512  
Db 1365 ITTDEVVTGLPI-----DLETGL-----PRDPVSGLPQLPNGTL 1397  
QY 513 TTGEVKKHLYGDNRYGGEPLFLPGEEGEE 541  
Db 1398 VDPSNKKPIFGS--HSG---FINGTSGEQ 1421

## RESULT 12

US-09-588-995A-5  
; Sequence 5, Application US/09588995A  
; Patent No. 6514697  
; GENERAL INFORMATION:  
; APPLICANT: PETERSEN, CAROLYN  
; APPLICANT: BARNES, DEBRA A.  
; APPLICANT: NELSON, RICHARD C.  
; APPLICANT: GUT, JIRI  
; TITLE OF INVENTION: METHODS FOR DETECTION OF CRYPTOSPORIDIUM SPECIES AND  
; TITLE OF INVENTION: ISOLATES AND FOR DIAGNOSIS OF CRYPTOSPORIDIUM  
; TITLE OF INVENTION: INFECTIONS  
; FILE REFERENCE: 480.19-5  
; CURRENT APPLICATION NUMBER: US/09/588,995A  
; CURRENT FILING DATE: 2000-06-06  
; PRIOR APPLICATION NUMBER: 08/827,171  
; PRIOR FILING DATE: 1997-03-27  
; PRIOR APPLICATION NUMBER: 08/928,361  
; PRIOR FILING DATE: 1997-09-12  
; PRIOR APPLICATION NUMBER: 08/700,651  
; PRIOR FILING DATE: 1996-08-14  
; PRIOR APPLICATION NUMBER: 08/415,751  
; PRIOR FILING DATE: 1995-04-03  
; NUMBER OF SEQ ID NOS: 115  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 1837  
; TYPE: PRT  
; ORGANISM: Cryptosporidium parvum  
US-09-588-995A-5

Query Match 3.7%; Score 116.5; DB 4; Length 1837;

Best Local Similarity 19.9%; Pred. No. 0.039;  
Matches 113; Conservative 74; Mismatches 175; Indels 207; Gaps 30;

QY	48	LNVSALHPAPPALHPFKPQSSNSPAI-----VVKPKAKESNTKQNNL-----	88
DB	985	LNPAATGMIPEGSLGPEQTFPEIEDGGHIPPVEAANAADKFKLGIPPSPVESIPEKQ	1044
QY	89	-FORAAAAAADAAGFLVSEKHLPLPKPTADPSPQIAGNEFAPNEOPVRLENLFPVGLKPD	147
DB	1045	KIDSISELMYDIBSGRLIGQVSKRPIGS-----IAGDLNPIKMTPTQTD-SVTGKPID	1097
QY	148	SIGKVYVRNGANPLHPEVTHGHHFDGDMVHAKVFEHGSASY--ACRFTOTNRFVQERQL	205
DB	1098	PTTGL-----PFPNP-TGH-----LNPNTNNNTMDSFAGAYKIYVNSGIKTDNVY	1142
QY	206	GRFV-----FPKAIHELHGHTGIARLMLFYARAAAGIVDPAHGTGVAN--	248
DB	1143	GLEFVEITGLPKOPVSDIPFNSTTGEL-----VDPSTGKPINNYT	1182
QY	249	AGLVYFNGRLIAMSDDLFPQVOITPNDL---KTVGRFPDGOLESTMAHPKVDPESG	305
DB	1183	AGIV-SKRGELPIEDS-----NGNLFDPSTKLPIDGNOL-----VNPETN	1223
QY	306	ELFALSVDVVSYPYLKYFRFSP-----DGTGSPDVEI--QLDQ-----P	342
DB	1224	STVSGSTSGTKP-----KGPVPVNGGVVPDEEAKDQADKDGKGLIIVPPTNSINKOP	1276
QY	343	TMHDFAITENFVVVPDQVFKLPEMIRG-----GSPVVDKKNVAREFGILDKYAE	395
DB	1277	VTHQVSNNTGNIINPTGKV--IPGSLPGLSNVPSNTTQOTDE-----ITGKPVDT	1327
QY	396	SSNIKWDIADPCFCFHLNNAWEPEPTDEVVVGISGCMTPPDSIFNESDENLKSVLSEIRLN	455
DB	1328	VTGLPY-----DPTGEIIDPATKLPISGVAGD-----EILTVEV-LN	1364
QY	456	LKGTSESTERPIISNEDQVNLAGEVMNRNMLGRKTKFAYLALAEPPWPKVSGFAKV--DL	512
DB	1365	ITDEVTLPI-----DLETGL-----PRDPVSGLPQLPNGTL	1397
QY	513	TTGEVKKHLYGDNRYGGEPLFLPGEGEE	541
DB	1398	VDPNKKPIPGS--HSG---FINGTSGEQ	1421

RESULT 13

US-08-311-731A-403

Sequence 403, Application US/08311731A

Patent No. 6583266

GENERAL INFORMATION:

APPLICANT: SMITH, DOUGLAS

APPLICANT: MAO, JEN-I

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES

TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR

TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 411

CORRESPONDENCE ADDRESS:

ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.

STREET: 600 ATLANTIC AVENUE

CITY: BOSTON

STATE: MASSACHUSETTS

COUNTRY: USA

ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/311,731A

FILING DATE:

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: GATES, EDWARD R.

REGISTRATION NUMBER: 31,616

REFERENCE/DOCKET NUMBER: C0044/7125

TELECOMMUNICATION INFORMATION:

Db 267 DAKPVKIKBEFVSIDI-----TFVSELEADLASGDSLPMGVILGAOSERFPSNLEVEAGP 322  
RESULT 15  
US-09-107-532A-5705  
; Sequence 5705, Application US/09/107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 5705:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 471 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...471  
; SEQUENCE DESCRIPTION: SEQ ID NO: 5705:  
US-09-107-532A-5705  
Query Match 3.3%; Score 105.5; DB 4; Length 471;  
Best Local Similarity 19.7%; Pred. No. 0.05; Indels 175; Gaps 27;  
Matches 97; Conservative 72; Mismatches 149;  
Qy 179 AVKEHGSASVACFTTNR---FVQERQLG-----RPVFPKAI-----GELHGH 220  
Db 47 SWIRFGDTTVCISATIENSVPFPLRETGTGWTAEYSMLPRATSTNRRESKGLFGR 106  
Qy 221 T-GIARLMLFYARAAGIVPAHGTGVANAGLVYFNGRLLAMSEDDLPLYOVQITPNDLK 279  
Db 107 TMEIQRLI---GRSLRAVVD-----LEKLGERSIIVDCD-----VIQADGGTR 146  
Qy 280 TVGRFDFDGLQLESTMIHPKV-----DPESGELFALSVDVVSKEYLKYFRFSPDGTK 331  
Db 147 TA---SITGAPVALKLALEKRELSDEPIKEHLAAVSVGIL-----PDGTC 192  
Qy 332 SPDVEIQLDQPTMMH-DFATEN--FVVV--PDQOVF---KLPEMIRGSPVYDKNKV'383

Db 193 VTDLDYQEDSAAALVDMNLVMTESGKFVEIQGTGBEATFDGSQLNEMLPFGKNAIEDLIKE 252  
Qy 384 ARFGILDKYABDSSNIKWIDAPDCFCFHLNNAWBEPEDETVVVGSCMTPPD-----SIF 438  
Db 253 OKHALLTEFAQNDERI-----SETKT-----IIATRNPGKAEFRNMF 291  
Qy 439 NESDENLKSVL-----SEIRLNKLT-GESTRPLISNED-QQVNLKXGM 480  
Db 292 KEAGYHVKTLLDYPPELDPVEETGTSTFEENARLKAETIAQLDDOPVLADDSGLKVDALGGM 351  
Qy 481 -----VNRNMLGRKTKPAY-----LALAE-----WPK 503  
Db 352 PGIIYSARFAGQKSDAGNNAKLLYELTDVPDEKTAQHCTLVFAAPKDSLVVEAEP- 410  
Qy 504 VSGPAKVDLTGTEVKKHLYGDNRYGGEPLFLPGEGGEDEGYILCFVHDEKTKWSELOIV 563  
Db 411 -----GRVARIPSGENGFGYDPLFIP-EGKQTAAEL-----SSEK----- 446  
Qy 564 NAVSLEVEATVKL 576  
Db 447 NKISHRAQAMKKL 459

Search completed: November 14, 2003, 21:18:46  
Job time : 23 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 21:11:32 ; Search time 47 Seconds  
(without alignments)  
2022.920 Million cell updates/sec

Title: US-09-758-269-6  
Perfect score: 3150  
Sequence: 1 MASFTATAVSGRWLGNGHT.....VPYGHGTFIGADDLAKQV 599

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

```
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3150	100.0	599	AAE04784	Arabidopsis thalia
2	3150	100.0	599	ABR92311	Herbicidally activ
3	2280.5	72.4	605	AAE04789	Lycopersicon escul
4	2280.5	72.4	605	AAE04789	Neoxanthin cleavag
5	2168.5	61.2	589	AAE04787	Vigna unguiculata
6	2044.5	64.9	589	ABR91182	Herbicidally activ
7	1991	63.2	583	AAE04782	Arabidopsis thalia
8	1930	61.3	604	AAE04788	Zea mays neoxanthi
9	1930	61.3	604	AAE04788	Neoxanthin cleavag

10	1663.5	52.8	577	22	AAE04786	Arabidopsis thalia
11	1663.5	52.8	577	23	ABR92416	Herbicidally activ
12	968	30.7	595	21	AAG31333	Arabidopsis thalia
13	968	30.7	595	22	AAE04783	Arabidopsis thalia
14	968	30.7	595	22	AAE04783	Neoxanthin cleavag
15	959	30.4	517	21	AAG31334	Arabidopsis thalia
16	939	29.8	538	21	AAG31316	Arabidopsis thalia
17	939	29.8	538	21	AAG31315	Arabidopsis thalia
18	938	29.8	538	22	AAE04785	Arabidopsis thalia
19	937	29.7	538	22	AAE04790	Arabidopsis thalia
20	925.5	29.4	501	21	AAG31317	Arabidopsis thalia
21	857.5	27.2	544	22	AAE04783	Sunflower neoxanth
22	834.5	26.5	431	21	AAG31335	Arabidopsis thalia
23	745	23.7	445	22	AAE04783	Neoxanthin cleavag
24	322.5	10.2	570	22	AAE12056	Arabidopsis thalia
25	307.5	9.8	596	22	AAE12085	Arabidopsis thalia
26	279	8.9	177	22	AAE09881	Physcomitrella pat
27	272.5	8.7	166	21	AAG32792	Zea mays protein f
28	264.5	8.4	153	21	AAG32793	Zea mays protein f
29	260	8.3	526	21	AAE04783	Beta, Beta-caroten
30	259.5	8.2	152	21	AAE04783	Beta, Beta-caroten
31	256	8.1	515	22	AAE04783	Amino acid sequenc
32	247	7.8	506	21	AAE04783	Beta, Beta-caroten
33	238.5	7.6	549	22	AAU04293	Zebra-2, Beta-caro
34	232.5	7.4	532	22	AAU04292	Mouse-2, Beta-caro
35	225	7.1	547	23	ABP63051	Human polyepitide
36	223	7.1	547	22	AAE03108	Human protein sequ
37	210	6.7	533	22	AAE03955	Canine RPE55. Can
38	209.5	6.7	533	22	AAE03955	Human protein sequ
39	209	6.6	533	14	AAE04617	Human retinol bind
40	207.5	6.6	556	22	AAU04294	Human-2, Beta-caro
41	206.5	6.6	579	22	AAE06872	Human RECAP polyep
42	185.5	5.9	529	21	AAE07315	Beta, Beta-caroten
43	154.5	4.9	65	23	ABP07713	Human ORFX protein
44	150	4.8	620	22	ABE071747	Drosophila melanog
45	150	4.8	620	22	AAU04291	Drosophila beta-ca

#### ALIGNMENTS

##### RESULT 1

AAE04784  
ID AAE04784 standard; Protein; 599 AA.

XX AC AAE04784;

XX DT 10-SEP-2001 (first entry)

XX DE Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED3.

XX KW Neoxanthin cleavage enzyme; AtNCED3; abscisic acid; ABA; herbicide;  
XX NW stress tolerance; transgenic plant; plant breeding; antisense-therapy;  
XX KW plant growth protectant.

XX OS Arabidopsis thaliana.

XX PN EP1116794-A2.

XX PD 18-JUL-2001.

XX PF 11-JAN-2001; 2001EP-0300218.

XX PR 13-JAN-2000; 2000JP-0010056.

XX PR 11-JAN-2001; 2001JP-0003476.

XX PA (RIKE ) RIKEN KK.

XX PI Iuchi S, Kobayashi M, Shinozaki K;

XX DR WPI; 2001-400081/43.

XX DR N-PSDB; AAD09396.

XX XX



PT	A DNA encoding a protein with a neoxanthin cleavage activity for
PPT	producing transgenic plants with improved or decreased stress tolerance
PT	-
XX	
PS	Claim 3; Fig 9; 10ipp; English.
PS	
XX	
CC	The invention relates to neoxanthin cleavage enzymes and their
CC	corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
CC	role in endogenous abscisic acid (ABA) biosynthesis under drought stresses
CC	Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC	plant when expressed in a plant cell. The invention also relates to
CC	methods for increasing or decreasing stress tolerance in a plant by
CC	introducing the DNA into the plant, and a transgenic plant into which a
CC	neoxanthin cleavage enzyme is introduced. The improvement of stress
CC	tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC	cleavage enzyme genes are useful for producing transgenic plants. An ar
CC	land can be improved by growing transformant weed for several years and
CC	then removing the weed by specifically lowering stress tolerance in the
CC	weed by inducing an inducible promoter. The present sequence is
CC	Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCE3 protein.
CC	The AtNCE3 thaliana neoxanthin cleavage enzyme, AtNCE3 protein.
CC	The AtNCE3 thaliana neoxanthin cleavage enzyme, AtNCE3 protein.
CC	library using a cDNA of the cFR65 (Cowpea Responsive to Dehydration)
CC	gene isolated from cowpea plant as a probe.
CC	
XX	
XX	Sequence 599 AA:
SO	

ID	AB92311 standard; Protein; 599 AA.
XX	
AC	AB92311;
XX	
DT	31-MAY-2002 (first entry)
XX	
DE	Herbicidally active polypeptide SEQ ID NO 1522.
XX	
KW	Herbicidal; plant; agriculture; herbicide.
XX	
OS	Arabidopsis thaliana.
XX	
PN	WO200210210-A2.
XX	
PD	07-FEB-2002.
XX	
PF	28-AUG-2001; 2001WO-EP09892.
XX	
PR	28-AUG-2001; 2001WO-EP09892.
XX	
PA	(FARB ) BAYER AG.
XX	
PI	Tietjen K, Weidler M;
XX	
WI	WPI; 2002-269010/31.
XX	
PT	Identifying plant target proteins for herbicidally active compounds,
XX	
PT	comprising aligning and comparing nucleic acid or amino acid sequences
XX	
PT	from plant with nucleic acid or amino acid sequences from non-plant
XX	
PS	organisms -
XX	
PS	Claim 5; SEQ ID NO 1522; 261pp + Sequence Listing; English.
XX	
CC	The invention relates to identifying target proteins
XX	
CC	(AB920790-AB924016) for herbicidally active compounds, comprising
XX	
CC	aligning and comparing nucleic acid or amino acid sequences from plant
XX	
CC	with nucleic acid or amino acid sequences from non-plant organisms using
XX	
CC	suitable search parameters, where plant sequences having an E-value
XX	
CC	greater by a factor of 3 than the E-value of most similar non-plant
XX	
CC	sequences are selected. The polypeptides or nucleic acids encoding them
XX	
CC	are useful for identifying modulators. The identified modulators are
XX	
CC	useful as herbicides.
XX	
SQ	Sequence 599 AA;
	Query Match 100.0%; Score 3150; DB 23; Length 599;
	Best Local Similarity 100.0%; Pred. No. 2e-310; Mismatches 0; Indels 0; Gaps 0;
	Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MASFTATAAASGRWLGGNHTQPPLSSSQSDLSYCSSLPMASRVTRKLNVSALHTPPAL 60
Db	1 MASFTATAAASGRWLGGNHTQPPLSSSQSDLSYCSSLPMASRVTRKLNVSALHTPPAL 60
QY	61 HFPQSSNSPAIVVKPKAESNTKQMWLCRRAAAALDAAGFLVSHSKLHPLKTDAPS 120
Db	61 HFPQSSNSPAIVVKPKAESNTKQMWLCRRAAAALDAAGFLVSHSKLHPLKTDAPS 120
QY	121 VOIAGTNAPVNEQPVRENLPVVGKLFDPSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAV 180
Db	121 VOIAGTNAPVNEQPVRENLPVVGKLFDPSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAV 180
QY	181 KPEHGSAACYACRFQTQNRVFQESQLGRPVPPKAIGELHGHGTGIARIMLFYAAAAAGIVDP 240
Db	181 KPEHGSAACYACRFQTQNRVFQESQLGRPVPPKAIGELHGHGTGIARIMLFYAAAAAGIVDP 240
QY	241 AHGTVANAGLVLYENGRLLAASDDLLPYOVQITPNGDLKTGCRFDFDQGLESTMAHPKV 300
Db	241 AHGTVANAGLVLYENGRLLAASDDLLPYOVQITPNGDLKTGCRFDFDQGLESTMAHPKV 300
QY	301 DPESGEIPALSVDVVSKPYLKYPFRSPDGTKSPDVEIQLDQPTMMHDFAITENFVVVPDQ 360
Db	301 DPESGEIPALSVDVVSKPYLKYPFRSPDGTKSPDVEIQLDQPTMMHDFAITENFVVVPDQ 360

QY 361 QVFKLPBEMIRGSPVYVDKVKARFGLDKYAEDESSNKKWIDAPDCFCFHLWNWEEPE 420  
 DB 361 QVFKLPBEMIRGSPVYVDKVKARFGLDKYAEDESSNKKWIDAPDCFCFHLWNWEEPE 420  
 QY 421 TDEWVIGSCWTPPDSIFNESDENLKSLSVLSSEIRLNKLTGSESTRPPIISNEDQOVNLEAGW 480  
 DB 421 TDEWVIGSCWTPPDSIFNESDENLKSLSVLSSEIRLNKLTGSESTRPPIISNEDQOVNLEAGW 480  
 QY 481 VNRNMLGRKTKFAYLALAEAPWPKVSGFAKVDLTGTEVKHLYGDNRYGGEPLPLPGEIGE 540  
 DB 481 VNRNMLGRKTKFAYLALAEAPWPKVSGFAKVDLTGTEVKHLYGDNRYGGEPLPLPGEIGE 540  
 QY 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSPRYPGFHTFGADDLAKQVV 599  
 DB 541 EDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSPRYPGFHTFGADDLAKQVV 599

## RESULT 3

AAE04789  
 ID AAE04789 standard; Protein; 605 AA.

XX AC AAE04789;

XX DT 10-SEP-2001 (first entry)

XX LYcopersicon esculentum neoxanthin cleavage enzyme, LeNCEd1.

XX KW Tomato; neoxanthin cleavage enzyme; LeNCEd1; abscisic acid; ABA;  
 stress tolerance; transgenic plant; plant breeding; antisense-therapy;  
 XX plant growth protectant; herbicide.

XX OS LYcopersicon esculentum.

XX PN EPI116794-A2.

XX PD 18-JUL-2001.

XX PF 11-JAN-2001; 2001EP-0300218.

XX PR 13-JAN-2000; 2000JP-0010056.

XX PR 11-JAN-2001; 2001JP-0003476.

XX PA (RIKE ) RIKEN KK.

XX PI Tsuchi S, Kobayashi M, Shinozaki K;

XX DR WPI; 2001-400081/43.

XX DR N-PSDB; AAD09401.

XX PT A DNA encoding a protein with a neoxanthin cleavage activity for  
 producing transgenic plants with improved or decreased stress tolerance

XX FS Claim 3; Fig 2; 101pp; English.

XX CC The invention relates to neoxanthin cleavage enzymes and their  
 corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key  
 role in endogenous abscisic acid (ABA) biosynthesis under drought stress.  
 CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a  
 plant when expressed in a plant cell. The invention also relates to  
 methods for increasing or decreasing stress tolerance in a plant by  
 introducing the DNA into the plant, and a transgenic plant into which a  
 neoxanthin cleavage enzyme is introduced. The improvement of stress  
 tolerance in plants is useful, for example in plant breeding. Neoxanthin  
 CC cleavage enzyme genes are useful for producing transgenic plants. An arid  
 land can be improved by growing transformant weed for several years and  
 then removing the weed by specifically lowering stress tolerance in the  
 CC weed by inducing an inducible promoter. The present sequence is  
 LYcopersicon esculentum neoxanthin cleavage enzyme, LeNCEd1 protein  
 CC related to the invention.

XX Sequence 605 AA;

Query Match 72.4%; Score 2280.5; DB 22; Length 605;  
 Best Local Similarity 71.7%; Pred. No. 4.6e-222;  
 Matches 435; Conservative 63; Mismatches 90; Indels 19; Gaps 6;  
 QY 5 TATAAAGRWGGNHGHTOPPLSSSSQSDLSYCS---SLPMASRVTRKLNVSALHTPPALH 61  
 DB 3 TTTSHATNTWI-----KTKLSMPSSKEFGFASNISLLKQHNQSLNINSSLOAPPILH 57  
 QY 62 PPKSSN--SPA--IVVKPKAKESN-----TKQNLIFQRAAAALDAAEFLVSEKHLHP 112  
 DB 58 PPKSSNYQTPEKNTISHPKQENNNSSSSSTSKNVLVQKAAAMALDAVESALTTHLEHP 117  
 QY 113 LPKTDAPSVQIAGNAPVNEOPVRRNLVWVKLPDSIKGVVVRNGANPLHEPVTGHHFFD 172  
 DB 118 LPKTDAPSVQIAGNAPVNEOPVRRNLVWVKLPDSIKGVVVRNGANPLHEPVTGHHFFD 177  
 QY 173 GDMVHAVKFEHGSASVACRFTQTNRFVQERQLGRPVFPKAIGBLHGHGTIARLMLFYAR 232  
 DB 178 GDMVHAVQFKNGSASVACRFTETETRLVQEKALGRPVFPKAIGBLHGHGTIARLMLFYAR 237  
 QY 233 AAGIVDPAHGTGVANAGLVFNGRLAMSDDLPYQVQITPNGDLTKTVGRDFDGOLES 292  
 DB 238 GLFLVDHSGKTGVANAGLVFNNRLAMSDDLPYHVKVPTGDLTKTEGRDFDGOLES 297  
 QY 293 TWIAHPKVDPSGELFALSYDWSKPYLKFRFSPDGTSPDVEIQLDQPTMMHDFATE 352  
 DB 298 TWIAHPKLDPVSGELFALSYDVIQPKYKFRFSGKNGESNDVPIPVEDPTMMHDFATE 357  
 QY 353 NFVVPDQOVVFKLPBEMIRGSPVYVDKVKARFGLDKYAEDESSNKKWIDAPDCFCFHL 412  
 DB 358 NFVVPDQOVVFKLPBEMIRGSPVYVDKVKARFGLDKYAEDESSNKKWIDAPDCFCFHL 417  
 QY 413 WNAMEEPTDEWVIGSCWTPPDSIFNESDENLKSLSVLSSEIRLNKLTGSESTRPPIISNEDQ 472  
 DB 418 WNAMEEPTDEWVIGSCWTPPDSIFNESDENLKSLSVLSSEIRLNKLTGSESTRPPIISNEDQ 477  
 QY 473 QVNLEAGMVNRNMLGRKTKFAYLALAEAPWPKVSGFAKVDLTGTEVKHLYGDNRYGGEPL 532  
 DB 478 QVNLEAGMVNRNMLGRKTKFAYLALAEAPWPKVSGFAKVDLTGTEVKHLYGDNRYGGEPL 537  
 QY 533 FLPGGE--GGEDEGYILCFVHDEKTKWSELQIVNAVSLVEATVKLPSPRYPGFHTFG 590  
 DB 538 FLPRDPSKEEDDGYILAFVHDEKTKWSELQIVNAVSLVEATVKLPSPRYPGFHTFG 597  
 QY 591 ADDLAKQ 597  
 DB 598 ANDLANQ 604

## RESULT 4

AAE72308  
 ID AAE72308 standard; Protein; 605 AA.

XX AC AAE72308;

XX DT 16-MAY-2001 (first entry)

XX DE Neoxanthin cleavage enzyme-like protein amino acid sequence.

XX KW Defence-related signalling gene; sunflower; neoxanthin cleavage enzyme;  
 NCE; amino acid permease; AAP; glutamic acid rich protein; GRP;  
 XX pathogen resistance; abscisic acid metabolism.

XX OS LYcopersicon esculentum.

XX PN WO200112801-A2.

XX PD 22-FEB-2001.

XX PF 17-AUG-2000; 2000WO-US22961.

XX PR 18-AUG-1999; 99US-0149656.

XX PR 23-MAY-2000; 2000US-0206405.







```
XX SQ Sequence 604 AA;
Query Match 61.3%; Score 1930; DB 22; Length 604;
Best Local Similarity 63.3%; Pred. No. 1.9e-186;
Matches 375; Conservative 73; Mismatches 122; Indels 22; Gaps 8;
QY 22 PFLSSQSSDLSYCSLPMASRVTRKLVNVSALHTP---PALHFPKSSNSPAIVV---- 74
Db 17 PARSARAS-----NSVRFSRAVSSVPPAECLQAPFHKPVADLPAPSKPAIAVPGHA 71
QY 75 -KPKAKESNTKOMNLQRAAAALDA-AEGFLVS-HEKHLPLKPTADPSVOIAGNAPVN 131
Db 72 AAPRKAEGGKKQLNLFQRAAAALDAFEFGVANVLERPHGLPSTADPAVQIAGNAPVG 131
QY 132 EQVVRNLVYVGGKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAKFEGHSA-SYA 190
Db 132 ERPPVHELPSGRIPPIDGVYARNGANPCFDPVAGHLLFDGGMVHALIRNGAAESYA 191
QY 191 CRFTQNRVFOEROLGRPFPPKAIAGELHGTGTARLMLFYARAAAGIVDPAGTGVANAG 250
Db 192 CRFTETARLQERAIQRPFVFPKAIAGELHGHSGTARLALFYARAAACGLVDPGAGTGVANAG 251
QY 251 LVYFNGRLLAMSEDDLPYQVQITPNGDLKTVGRFDFDGOLESTMIAPKVPDPESGELPAL 310
Db 252 LVYFNGRLLAMSEDDLPYHVRVADGDGLETVGRYDFDGOGLGCAMIAHPKLDPATGELHAL 311
QY 311 SYDVVSKPYLKYFRFSDGTGKSPDVEIQDPTMMDHFAITENFVVVPPQVQVFKLPEMI 370
Db 312 SYDVIKRPYLKYFYFRPDGTGKSDVEIPLQPTMIDHFAITENFVVVPPHVVQVFKLQEM 371
QY 371 RGGSPVVYDKNKVARFGILDKYAESDSNKKWIDAPDCFCFHLNNAWEEPEDEVVVGSC 430
Db 372 RGGSPVVLDEKTSRFGVLPKHAADASEMAWVDVPCFCFHLNNAWEEPEDEVVVGSC 431
QY 431 MTPPDSIFNEDSNLKSVLSEIRLNLTGSESTRRPIISNEDQOVNLEAGVMNRNMLGRKT 490
Db 432 MTPADSIFNEDSRLESVLTEIRLDARTGSTRRAVLP-PSQENLEVGMVNRNLLGRES 490
QY 491 KFYALALAEPPWPKVSGFAKVDLTGTVGVKKHLYGDNRYGGEPLFLPGEGGE-----EDEGY 545
Db 491 RYAYLAVAPWPKESGFAKEDLTGELTFEYGEGRFGGECFCFVPMDDAAHPREDGGY 550
QY 546 ILCFVHDEKTKWSELOIVNAVSLVEBATVKLPSRPYPYFGHGTFIGADDLIAQ 597
Db 551 VLTFFVHDERAGTSSELLVNAADIRLEATVQLPSRVFPFGHGTFIGTQOELEAQ 602
RESULT 9
AAB72309
ID AAB72309 standard; Protein; 604 AA.
XX
AC AAB72309;
XX
XX
DT 16-MAY-2001 (first entry)
XX
DE Neoxanthin cleavage enzyme-like protein amino acid sequence.
XX
KW Defence-related signalling gene; sunflower; neoxanthin cleavage enzyme;
KW NCE; amino acid permease; AAP; glutamic acid rich protein; GRP;
KW pathogen resistance; abscisic acid metabolism.
OS Zea mays.
XX
FN W0200112801-A2.
XX
XX
PD 22-FEB-2001.
XX
PF 17-AUG-2000; 2000WO-US22961.
XX
PR 18-AUG-1999; 99US-0149656.
XX
PR 23-MAY-2000; 2000US-0206405.
XX
```

PA (PION-) PIONEER HI-BRED INT INC.  
XX (CURA-) CURAGEN CORP.

PI Bidney DL, Crasta OR, Hu X, Lu G;

XX WPI; 2001-211215/21.

XX Novel isolated defence-related signalling gene isolated from sunflower  
PT encoding neoxanthin cleavage enzyme, amino acid permease or glutamic  
PT acid-rich protein useful for increasing resistance of plant to a  
PT pathogen

XX Disclosure; Fig 1; 135pp; English.

XX This invention relates to defence-related signalling genes isolated from  
the sunflower (*Helianthus annuus*). The genes encode a neoxanthin cleavage  
enzyme (NCE), an amino acid permease (AAP) and a glutamic acid rich  
protein (GRP). The signalling gene is useful for increasing the  
resistance of a plant to a pathogen such as fungus, virus, bacterium,  
nematode or insect (e.g. European corn borer), preferably  
Sclerotinia spp., Phoma spp., or Phomopsis spp., by stably incorporating a  
construct containing the gene into the genome of the plant. The gene is  
useful for regulating gene expression in a plant, in response to a  
stimulus such as infection with a pathogen, damage from a pathogen,  
hydrogen peroxide, jasmonic acid, methyl jasmonate, salicylic acid,  
oxalic acid or expression of a gene encoding oxalic acid oxidase. The  
genes are also useful for stem-preferred regulation of gene expression in  
a plant. The genes are useful in agriculture, particularly in the  
breeding of crop plants with improved agronomic traits, for modifying  
abscisic acid (ABA) metabolism and for modifying amino acid transport and  
content in plants. The present sequence represents a neoxanthin cleavage  
enzyme-like protein from *Zea mays* used in the characterisation of  
sunflower NCE.

XX Sequence 604 AA;

Query Match 61.3%; Score 1930; DB 22; Length 604;

Best Local Similarity 63.3%; Pred. No. 1.9e-186;

Matches 375; Conservative 73; Mismatches 122; Indels 22; Gaps 8;

QY 22 PFLSSQSSDLSYCSLPMASRVTRKLVNVSALHTP---PALHFPKSSNSPAIVV---- 74

Db 17 PARSARAS-----NSVRFSRAVSSVPPAECLQAPFHKPVADLPAPSKPAIAVPGHA 71

QY 75 -KPKAKESNTKOMNLQRAAAALDA-AEGFLVS-HEKHLPLKPTADPSVOIAGNAPVN 131

Db 72 AAPRKAEGGKKQLNLFQRAAAALDAFEFGVANVLERPHGLPSTADPAVQIAGNAPVG 131

QY 132 EQVVRNLVYVGGKLPDSIKGVYVRNGANPLHEPVTGHHFFDGDGMVHAKFEGHSA-SYA 190

Db 132 ERPPVHELPSGRIPPIDGVYARNGANPCFDPVAGHLLFDGGMVHALIRNGAAESYA 191

QY 191 CRFTQNRVFOEROLGRPFPPKAIAGELHGTGTARLMLFYARAAAGIVDPAGTGVANAG 250

Db 192 CRFTETARLQERAIQRPFVFPKAIAGELHGHSGTARLALFYARAAACGLVDPGAGTGVANAG 251

QY 251 LVYFNGRLLAMSEDDLPYQVQITPNGDLKTVGRFDFDGOLESTMIAPKVPDPESGELPAL 310

Db 252 LVYFNGRLLAMSEDDLPYHVRVADGDGLETVGRYDFDGOGLGCAMIAHPKLDPATGELHAL 311

QY 311 SYDVVSKPYLKYFRFSDGTGKSPDVEIQDPTMMDHFAITENFVVVPPQVQVFKLPEMI 370

Db 312 SYDVIKRPYLKYFYFRPDGTGKSDVEIPLQPTMIDHFAITENFVVVPPHVVQVFKLQEM 371

QY 371 RGGSPVVYDKNKVARFGILDKYAESDSNKKWIDAPDCFCFHLNNAWEEPEDEVVVGSC 430

Db 372 RGGSPVVLDEKTSRFGVLPKHAADASEMAWVDVPCFCFHLNNAWEEPEDEVVVGSC 431

QY 431 MTPPDSIFNEDSNLKSVLSEIRLNLTGSESTRRPIISNEDQOVNLEAGVMNRNMLGRKT 490

Db 432 MTPADSIFNEDSRLESVLTEIRLDARTGSTRRAVLP-PSQENLEVGMVNRNLLGRES 490

QY 491 KFYALALAEPPWPKVSGFAKVDLTGTVGVKKHLYGDNRYGGEPLFLPGEGGE-----EDEGY 545

```

Db 491 RYAYLVAEPWPKESGFAKEDLSTGELTKEFYGEFGCEPCVPMDPAAAHPRGEDDY 550
Qy 546 ILCFVHDEKTKWSELOIVNAVSLVEATVKLPRVPYGHGTGFIAGDDIAKQ 597
Db 551 VLFVHDERAGTSELLVNAADIRLEATVOLPSRVPFGHGTFTITGOELEAQ 602

```

## RESULT 10

AAE04786 standard; Protein; 577 AA.

AC AAE04786;

XX 10-SEP-2001 (first entry)

XX Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED5.

XX Neoxanthin cleavage enzyme; AtNCED5; abscisic acid; ABA; herbicide;  
 KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;  
 KW plant growth protectant.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

XX EP1116794-A2.

XX 18-JUL-2001.

XX 11-JAN-2001; 2001EP-0300218.

XX 13-JAN-2000; 2000JP-0010056.

XX 11-JAN-2001; 2001JP-0003476.

XX (RIKE ) RIKEN KK.

XX Iuchi S, Kobayashi M, Shinozaki K;

XX WPI; 2001-400081/43.

XX N-PSDB; AAD03398.

XX A DNA encoding a protein with a neoxanthin cleavage activity for  
 PT producing transgenic plants with improved or decreased stress tolerance

XX Claim 3; Fig 10; 101pp; English.

XX The invention relates to neoxanthin cleavage enzymes and their  
 CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key  
 CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.  
 CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a  
 CC plant when expressed in a plant cell. The invention also relates to  
 CC methods for increasing or decreasing stress tolerance in a plant by  
 CC introducing the DNA into the plant, and a transgenic plant into which a  
 CC neoxanthin cleavage enzyme is introduced. The improvement of stress  
 CC tolerance in plants is useful, for example in plant breeding. Neoxanthin  
 CC cleavage enzyme genes are useful for producing transgenic plants. An arid  
 CC land can be improved by growing transformant weed for several years and  
 CC then removing the weed by specifically lowering stress tolerance in the  
 CC weed by inducing an inducible promoter. The present sequence is  
 CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED5 protein.  
 CC The AtNCED5 cDNA is obtained from an Arabidopsis plant-derived cDNA  
 CC library using a cDNA of the CRD65 (CowPea Responsive to Dehydration)  
 CC gene isolated from cowpea plant as a probe.

XX Sequence 577 AA;

XX Query Match 52.8%; Score 1663.5; DB 22; Length 577;

XX Best Local Similarity 55.4%; Pred. No. 2.1e-159;

XX Matches 323; Conservative 100; Mismatches 137; Indels 23; Gaps 9;

XX 24 LSSQSSDLSYSCSLPMASRVTKLVSS-ALHTPPALHFPKQSSNSPAIVKPKAKESN'82

XX 9 LLPTKTSRSHLLQPKKNAVISRILLINPFKIPTLPDITSVPSP-----VKLKPTYPN 62

XX Db .

```

Qy 83 TKQNNLQRAAAALDAAE-GFLVSHKHLPLKPTADPSVQIAGNFAVNEQPVRRNLPV 141
Db 63 ---NLNLOKLAATWLDKIESIVIPMEQNRPLPAPTDFAVQLSGNFAVNECPVQNGLEV 119
Qy 142 VGKLPDSIKGVYVANGANPLHEPVTGHHFFDGDGMHAVK--FEHGSASACRFTQTNR 199
Db 120 VQIIPSCCLKGVYIRNGANPMFPPLAGHHLFDGDGMHAVSIGFDN-QVSYSCTYKTNRL 178
Qy 200 VOERQLGRPYPPKAIHELHGHGTGIALMLFYARAAAGIVDPAKGTGVANAGLVYFNGRLL 259
Db 179 VOETALGRSVFPKPIGELHGHSGLARLALFTARAGIGLVDTGRMGVANAGVVFNGRLL 238
Qy 260 AMSEDDLPYQVQIITPENGDLTKTVGRFDFDGOLESTMIAPKVPDPESGELFALSVDVWSKY 319
Db 239 AMSEDDLPYQVKIDGOGLDETIGRFDFDQDIDSSVIAHPKVDAITGDLHTLSYNLKKPH 298
Qy 320 LKYRFPDGTGKSPDVEIQDQPTMHDFAITENFVVVDDQVVKLPKEMIRGSPVYVD 379
Db 299 LRYLKFNTCGKKTRDVEITLPEPTMIHDAITENFVVIPOQMVFKLSEIRGSPVIYV 358
Qy 380 KMKVARFGILDKYAEDSNKWDAPDCFCFHLNMAWEEPETDE---VVVIGSCMTTPPD 435
Db 359 KEKVARGVLSKQDLTGSDINWVDVDFCFHLNMAWEE-RTTEGDFPVIIVIGSCMSPPD 417
Qy 436 SIFNESDENLKSVLSEIRLNKTCGSTRRPIISNEDQOVNLEAGMVRNMLGRKTKAYL 495
Db 418 TIFSEGEPTRVLSLSEIRLNKTCGSTRRPIISNEDQOVNLEAGMVRNMLGRKTKAYL 473
Qy 496 ALAPFPKPVSGFAKVDLTGTVKVKHLYGDNRYGCEPLFLPGEGEDEDEGYILCFVHDEKT 555
Db 474 AIADFPKPCSGIAKVDIQNGTVSEFNYPGSPFGGCEPCFVPEGEDEDEKGYVMGFVRDEK 533
Qy 556 WKSELOIVNAVSLVEATVKLPSRVPYGHGTFTIGADDLAKQV 598
Db 534 DESEFVVVDATDMKQAAVRLPERVPYGHGTFTVSENQLKEQV 576

```

## RESULT 11

ABB92416

ID ABB92416 standard; Protein; 577 AA.

XX AC ABB92416;

XX 31-MAY-2002 (first entry)

XX Herbicidally active polypeptide SEQ ID NO 1627.

XX Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.

XX WO200210210-A2.

XX 07-FEB-2002.

XX 28-AUG-2001; 2001WO-EP09892.

XX 28-AUG-2001; 2001WO-EP09892.

XX (FARB ) BAYER AG.

XX Tietjen K, Weidler M;

XX WPI; 2002-269010/31.

XX Identifying plant target proteins for herbicidally active compounds,  
 PT comprising aligning and comparing nucleic acid or amino acid sequences  
 PT from plant with nucleic acid or amino acid sequences from non-plant  
 PT organisms -

XX Claim 5; SEQ ID NO 1627; 261pp + Sequence Listing; English.

XX



CC The invention relates to identifying target proteins  
CC (AB90790-AB94016) for herbicidally active compounds, comprising  
CC aligning and comparing nucleic acid or amino acid sequences from plant  
CC with nucleic acid or amino acid sequences from non-plant organisms using  
CC suitable search parameters, where plant sequences having an E-value  
CC greater by a factor of 3 than the E-value of most similar non-plant  
CC sequences are selected. The polypeptides or nucleic acids encoding them  
CC are useful for identifying modulators. The identified modulators are  
CC useful as herbicides.  
XX  
XX Sequence 577 AA;  
  
Query Match 52.8%; Score 1663.5; DB 23; Length 577;  
Best Local Similarity 55.4%; Pred. No. 2.1e-159;  
Matches 323; Conservative 100; Mismatches 137; Indels 23; Gaps 9;  
  
QY 24 LSSQSQSDLSYCSLSPWASRVTRKLVSS-ALHTPPALHPKQSSNSPAIVVVKPKAKSN 82  
DB 9 LPTKTSRSHLLPQPKANISRIILNFKIPTLPDLTSFVPSP-----VKLKPTVFN 62  
  
QY 83 TKOMLQRAAAALDAAE-GFLVSHKHLHPKTAADPSVQIAGNFAPVNEQVVRNLPV 141  
DB 63 ---LNLQKLAATMLQKIESSIVIPMEQNRPLPKPTDPAVQLSGNFAPVNECPVQNGLEV 119  
  
QY 142 VGLPDSIKGVYRNGANPLHEPVTCGHHFDDGGMVHAK--PEHGSASYACRPTOTRPF 199  
DB 120 VGIPLSCLGVYIRNGANPMFPPLAGHHLFDGDMIHAVSIGFDN-QVSYSCRYTKTRL 178  
  
QY 200 VQRQLGRVPFPAKIGELGHGHTGIARLMFLYARAAAGIVDPAHGTGVANAGLVYFNGRL 259  
DB 179 VQETALGRSVFPKPIGELGHGSLARLAFTRAGLIGVDGTRGMGVANAGVVFNGRL 238  
  
QY 260 AMSEDLPLQVQITPNDGKTVGRFDFDQLSSTMIHAKVPDSEGLFPALSYDVVSKPY 319  
DB 239 AMSEDLPLQVQITDGGDLDTIGRFQDDQIDSSVIAHPKVDATTGDLTSLYNVLKPKH 298  
  
QY 320 LKVFSPDGTGKSPDVEIOLDPTMMHDAITENFVVDPDQVFKLPEMIRGSGPVVD 379  
DB 299 LRLKXNTGKTRDVEITLPTMTLHDAITENFVDPQVFKLSEMRGSGPVIV 358  
  
QY 380 KKKVAFGLDKYAEVDSNKKIMADPCFCFLHNAWESPEPTE----VVVIGSCMTPPD 435  
DB 359 KKKVAFGLDKYAEVDSNKKIMADPCFCFLHNAWESPEPTE----VVVIGSCMTPPD 417  
  
QY 436 STFNESDENLKSVLSEIRLNLTGSTRPILSNEDQVNLKAGVNNLGRKTKAYL 495  
DB 418 TIFSEGEPTRVLSIRLNLTGSTRPILSNEDQVNLKAGVNNLGRKTKAYL 473  
  
QY 496 ALAEPKVSFGAKVDLTGTEVKKHLYGDNRYGGEPLFLPGGGEDEGYILCFVHDEKT 555  
DB 474 ATADPWFKSGIAKVDIQNGTVSEFNYPGSRFGGEPFVPEGEDEGKYVMGFVRDEK 533  
  
QY 556 WKSELQIVNAVSLVEATVKLPSRVPYGFHGTGFIAGDADLAKOV 598  
DB 534 DESEFVVVDATDMKQAAVRLPERVPGFHGTGFIAGDADLAKOV 576  
  
RESULT 12  
AAG31333  
ID AAG31333 standard; Protein; 595 AA.  
XX  
AC AAG31333;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37612.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX

PN EP1033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 18-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139839.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.

PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0156659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0158293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144332.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145115.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160981.
PR	26-JUL-1999;	99US-0145275.	PR	22-OCT-1999;	99US-0160989.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145919.	PR	25-OCT-1999;	99US-0161406.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161359.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146388.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161920.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161992.
PR	04-AUG-1999;	99US-0147204.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147302.	PR	29-OCT-1999;	99US-0162142.
PR	05-AUG-1999;	99US-0147192.			
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149902.			
PR	23-AUG-1999;	99US-0149930.			
PR	25-AUG-1999;	99US-0150566.			
PR	26-AUG-1999;	99US-0150884.			
PR	27-AUG-1999;	99US-0151065.			
PR	27-AUG-1999;	99US-0151066.			
PR	27-AUG-1999;	99US-0151080.			
PR	30-AUG-1999;	99US-0151303.			
PR	31-AUG-1999;	99US-0151438.			
PR	01-SEP-1999;	99US-0151930.			
PR	07-SEP-1999;	99US-0152363.			
PR	10-SEP-1999;	99US-0153070.			
PR	13-SEP-1999;	99US-0153478.			
PR	15-SEP-1999;	99US-0154018.			
PR	16-SEP-1999;	99US-0154039.			
Qy	42	SRVTRKLVSSAHTPPALHFPKQSSNSPAIVVVKPKAKESNTKQNLQFRAAAALDAAE	101	42	SRVTRKLVSSAHTPPALHFPKQSSNSPAIVVVKPKAKESNTKQNLQFRAAAALDAAE
Db	42	SPITNPSDNNDRNKPKTLH---NRTNHTLVSSPKLRPEMTLATALF---TTVEDVIN	94	42	SPITNPSDNNDRNKPKTLH---NRTNHTLVSSPKLRPEMTLATALF---TTVEDVIN
Qy	102	GFLVSHEKULPLKPTADPSVQIAGNFAPVNEQPVERNLPVV-GKLPDSIKGVYVNGANP	160	102	GFLVSHEKULPLKPTADPSVQIAGNFAPVNEQPVERNLPVV-GKLPDSIKGVYVNGANP
Db	95	TFIDP-----PSPSVDPKHLSDNFAPVLDELPTDCEIINGTLPLSLNGYIINGNP	149	95	TFIDP-----PSPSVDPKHLSDNFAPVLDELPTDCEIINGTLPLSLNGYIINGNP
Qy	161	LHEPVTGHEFFDGDGMVHAKPEHGSASYACRFTQTNRFQVQERQGRPVFPKAIQELHGH	220	161	LHEPVTGHEFFDGDGMVHAKPEHGSASYACRFTQTNRFQVQERQGRPVFPKAIQELHGH
Db	150	QFLPRGPFYHLFDGDMHAIKIHNGKATLCSRYVVTYKYNVEKQTGAPVMPNVFSGFNGV	209	150	QFLPRGPFYHLFDGDMHAIKIHNGKATLCSRYVVTYKYNVEKQTGAPVMPNVFSGFNGV
Qy	221	T-GIARMLFYARAAAGIVDPAHGTGVANAGLVYNGRLLAMSEDDLPYQVQITNGDLK	279	221	T-GIARMLFYARAAAGIVDPAHGTGVANAGLVYNGRLLAMSEDDLPYQVQITNGDLK
Db	210	TASVARGALTAARVLGTQYNPVNGIGLANTSIAFFSNLFLGSDLPYAVRLTSGDUE	269	210	TASVARGALTAARVLGTQYNPVNGIGLANTSIAFFSNLFLGSDLPYAVRLTSGDUE
Qy	280	TVGRFDFDGOLESTMLAHPKVDPESEGLFALSVDVVSFKYKFRFSPDGTSKSPDVEI-Q	338	280	TVGRFDFDGOLESTMLAHPKVDPESEGLFALSVDVVSFKYKFRFSPDGTSKSPDVEI-Q
Db	270	TIGRYDFDGLKAMSMTAHPKTDIITGETAFRYGPV-PPFLTYFRFDSAGKQROVPIFS	328	270	TIGRYDFDGLKAMSMTAHPKTDIITGETAFRYGPV-PPFLTYFRFDSAGKQROVPIFS
Qy	339	LDQPTWMDPATEFNVVVDQOVFK---LPEMIRGSGPVVDKKNKVARFGILDKYABD	395	339	LDQPTWMDPATEFNVVVDQOVFK---LPEMIRGSGPVVDKKNKVARFGILDKYABD
Db	329	MTSPSFLHDFAITKRHAFABIQOLGMRMNNLDLVLEGGSPVGTGDKTGRPGVIFRYAGD	388	329	MTSPSFLHDFAITKRHAFABIQOLGMRMNNLDLVLEGGSPVGTGDKTGRPGVIFRYAGD
Qy	396	SSNKKWIDAPDCFCFELWNAWEEPEDEVVVTGSCMTPPDSIFNESDENKSLVLSIRLN	455	396	SSNKKWIDAPDCFCFELWNAWEEPEDEVVVTGSCMTPPDSIFNESDENKSLVLSIRLN
Db	389	ESEKMFVPGFNIIHAINADEDDGNSVLLAPNIMSIEHTLERMD-LVHALVEKVKID	447	389	ESEKMFVPGFNIIHAINADEDDGNSVLLAPNIMSIEHTLERMD-LVHALVEKVKID

Query Match 30.7%; Score 968; DB 21; Length 595;

Best Local Similarity 36.7%; Pred. No. 1.1e-28; Indels 36; Gaps 13;

Matches 210; Conservative 106; Mismatches 220;

QY 456 LKTGSTRPPIISNEDQQVNLEAGWVRNMLGRKTKFAYLALAEPPKVSFKAQVDLTG 515  
 DB 448 LVTGIVRRHPISAR-----NLDFAVINPAFLGRCSRYVIAAGIDPMPKISGVKLDVSKG 502  
 QY 516 E-----VKKHLGDNRYGGEPLFLPGGG-----EDEGYILCFVHDEKTKWSELQIVNAV 566  
 DB 503 DRDDCTVARRMYGCGYGGEPFFVARDPGNPEAEEDDGYVTVYVHDEVTSKFLVMDAK 562  
 QY 567 S--LEVEATVKLPSRVPGHGHTIGADDLAK 596  
 DB 563 SPELEIVAARVLRPRVPYGFHGLFVKESDLNK 594

## RESULT 13

AAE04783  
 ID AAE04783 standard; Protein; 595 AA.

XX AAE04783;

XX 10-SEP-2001 (first entry)

DE Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED2.

XX Neoxanthin cleavage enzyme; AtNCED2; abscisic acid; ABA; herbicide;  
 KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;  
 KW plant growth protectant.

XX Arabidopsis thaliana.

XX EP1116794-A2.

XX 18-JUL-2001.

XX 11-JAN-2001; 2001EP-0300218.

XX 13-JAN-2000; 2000JP-0010056.

XX 11-JAN-2001; 2001JP-0003476.

XX (RIKE) RIKEN KK.

XX Iuchi S, Kobayashi M, Shinozaki K;

XX WPI; 2001-400081/43.

XX N-PSDB; AAD09395.

PT A DNA encoding a protein with a neoxanthin cleavage activity for  
 PT producing transgenic plants with improved or decreased stress tolerance

PT -

PS Example 10; Fig 10; 101pp; English.

XX The invention relates to neoxanthin cleavage enzymes and their  
 CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key  
 CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.  
 CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a  
 CC plant when expressed in a plant cell. The invention also relates to  
 CC methods for increasing or decreasing stress tolerance in a plant by  
 CC introducing the DNA into the plant, and a transgenic plant into which a  
 CC neoxanthin cleavage enzyme is introduced. The improvement of stress  
 CC tolerance in plants is useful, for example in plant breeding. Neoxanthin  
 CC cleavage enzyme genes are useful for producing transgenic plants. An arid  
 CC land can be improved by growing transformant weed for several years and  
 CC then removing the weed by specifically lowering stress tolerance in the  
 CC weed by inducing an inducible promoter. The present sequence is  
 CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED2 protein.  
 CC The AtNCED2 cDNA is obtained from an Arabidopsis plant-derived cDNA  
 CC library using a cDNA of the CPR65 (CowPea Responsive to Dehydration)  
 CC gene isolated from cowpea plant as a probe.

XX Sequence 595 AA;

XX Query Match 30.7%; Score 968; DB 22; Length 595;

XX Best Local Similarity 36.7%; Pred. No. 1.1e-88;

Matches 210; Conservative 106; Mismatches 220; Indels 36; Gaps 13;  
 QY 42 SAVTRKLVNSALHTPPALHFPKQSSNSPAIVVFPKAKESNTKQNVLFQRAAAALDAE 101  
 DB 42 SPITNPSNNRRNKPKTLH---NRNHTLVSSPPKLRPEMTLALALF---TTVEDVIN 94  
 QY 102 GFLVSHKHLPLKPTADPSVQIAGNFAPVNPQPVRRNLPPVV-GKLPDSIKGYVVENGANP 160  
 DB 95 THIDP-----PSRPSVDPKHVLSDNFAFVLDLPDCEIIHGTLPLSLNGAYIRNGNP 149  
 QY 161 LHEPVTGHHFFDGDGMVHAVKFEHGSASYACRFQTNRFOERQGLGRVFPFKAIGELHGH 220  
 DB 150 QFLPRGPYHLFDGDGMHLAIKHNGKATLCRSYKTYKYNVEKQTGAFWNPVFSFGNGV 209  
 QY 221 T-GIARLMLFYAAAAAGIVDPAHGTGVANAGLVYFNGRLLAMSEDDLPYQVQITPNGDLK 279  
 DB 210 TASVARGALTAARVLTGVNPNVNGIGLANTSLAFFSNELFALGESDLFAYVRLTESGDIE 269  
 QY 280 TVGRFDGQLESTMIAPKVDPESGELFALSYDVVSKPYLYKFRFSPDGTGKSPDVEI-Q 338  
 DB 270 TIGRYDFDGLKAMSMTAHPKTDPTITGETFAFRYGFV-PPFLTYFRFDSAGKKQRDPVIFS 328  
 QY 339 LDQPTMMHDFALTENFVVVDQVVFK---LPEMIRGGSPVVYDKNVARFGILDKYAE 395  
 DB 329 MTSPFLHDFAITKHAIFAETQLGRNMMLDLVLEGGSPVGTGDKTFLRGVIEKYAGD 388  
 QY 396 SSNIKWIDAPDCFCFLNMAWEEPETDEVVIGSCWTPPDSIFNESDENLKSVLSEIRLN 455  
 DB 389 ESEMKWFVEVGFNIIHAINAWDEDDGNSVLIAPNINSIEHTLERMD-LVHALVEKVKID 447  
 QY 456 LKTGSTREPIISNEDQQVNLEAGWVRNMLGRKTKFAYLALAEPPKVSFKAQVDLTG 515  
 DB 448 LVTGIVRRHPISAR-----NLDFAVINPAFLGRCSRYVIAAGIDPMPKISGVKLDVSKG 502  
 QY 516 E-----VKKHLGDNRYGGEPLFLPGGG-----EDEGYILCFVHDEKTKWSELQIVNAV 566  
 DB 503 DRDDCTVARRMYGCGYGGEPFFVARDPGNPEAEEDDGYVTVYVHDEVTSKFLVMDAK 562  
 QY 567 S--LEVEATVKLPSRVPGHGHTIGADDLAK 596  
 DB 563 SPELEIVAARVLRPRVPYGFHGLFVKESDLNK 594

## RESULT 14

AAE72306

ID AAE72306 standard; Protein; 595 AA.

XX AAE72306;

XX 16-MAY-2001 (first entry)

DE Neoxanthin cleavage enzyme-like protein amino acid sequence.

XX Defence-related signalling gene; sunflower; neoxanthin cleavage enzyme;  
 KW NCE; amino acid permease; AAP; glutamic acid rich protein; GAP;  
 KW pathogen resistance; abscisic acid metabolism.

OS Arabidopsis thaliana.

XX WO200112801-A2.

XX 22-FEB-2001.

XX 17-AUG-2000; 2000WO-US22961.

XX 18-AUG-1999; 99US-0149656.

XX 23-MAY-2000; 2000US-0206405.

XX (PION-) PIONEER HI-BRED INT INC.

XX (CURA-) CURAGEN CORP.

XX Bidney DL, Crasta OR, Hu X, Lu G;



```
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0143277.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.

PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 30.4%; Score 959; DB 21; Length 517;
Best Local Similarity 39.2%; Pred. No. 6.8e-88;
Matches 197; Conservative 98; Mismatches 183; Indels 24; Gaps 10;

QY 112 PLPKTADPSQIAGNAPAPNEQSVRENLFVV-GKLPDSIKGVYVRNGANPLHEPVTGHHF 170
Db 22 PSRPSVDPKRVLSNDNPAVLDELPTDCEIIHGTLPLSNGAYIRNGPNQFLPRGPYHL 81
QY 171 FDGDMVHAVKFEHGSASYACRFTQTNRFVQERQGRPVFPKPAIGELHGT-GIARLMLF 229
Db 82 FDGDGMLHAIKHNGKATLCSRYVKTKYNEKQTCAPVMPNVFSGFNGVTASVARGALT 141
QY 230 YABAAAGIVDPAHGTGVANAGLVYFNGRLLANSEDDLPYQVQITNGDLKTVGRDFDQG 289
```

